Comparing measures of stressful life events in young people

Brown, Hannah Mary

Awarding institution:
King's College London

Volume 1: Systematic Review and Main Research Project

Hannah Brown

Department of Psychology,

Institute of Psychiatry, Psychology & Neuroscience

King's College London

Thesis submitted in partial fulfilment for the degree of Doctorate in Clinical Psychology 2018

Acknowledgements

First, I would like to give special thanks to my supervisors, Dr Eleanor Leigh and Dr Patrick Smith for their knowledge, encouragement and feedback.

Second, many thanks go to all those involved in this study, including all the young people who took the time to complete the questionnaires and shared their experiences, the teachers and school staff who accommodated our research, and to all the research assistants who helped with data collection and entry.

Finally, on a personal note, thank you to my family and friends who supported me throughout the doctorate. I would not have been able to do this without you all. Special thanks to my husband and son for tolerating the hours spent absent from family life whilst writing, and thanks to my mum and dad for proof-reading another thesis! I promise this is the last one!

A systematic review of measures of stressful life events in young people

Abstract

Stressful events in the lives of young people have been measured using many methods. The diversity of measures makes it difficult to draw conclusions across studies regarding the role these events play in the development of common mental health problems in young people. The current review appraises measures of stressful life events in young people and assesses the quality of the methods used to develop such measures. Following PRISMA guidelines, published studies reporting on the psychometric properties of measures of stressful life events used with young people 18 years or younger were captured using systematic search terms across three databases. Study quality was assessed using the Consensus-based standards for the Selection of Health Measurement Instruments (COSMIN) scale. Psychometric properties (content validity, internal consistency, test-retest reliability, structural validity, criterion validity, cross-cultural validity) were assessed. In total, 21 studies were included describing unique life event measures. The majority referred to self-report checklists used in middle childhood. Nineteen studies examined content validity; 12 assessed internal consistency and 10 assessed test-retest reliability. Few studies examined the structural, criterion and cross-cultural validity of life event measures. The current review highlights the diversity of measures employed in the field and inconsistency in the methodological rigour with which they are developed. Recommendations about the use of measures of stressful life events and future research are provided.

Introduction

"Few constructs in mental health and psychopathology have been as important, yet at the same time as difficult to define, as the concept of "stress."

Grant et al. (2003)

Conceptualising stressful life events

Stressful life events are broadly defined as environmental circumstances or conditions that cause individuals psychological or biological distress. However, the construct of stressful life events has been conceptualised in several different ways.

Compas (1987) provides a useful illustration for understanding different conceptualisations of stressful life events (Figure 1). Whilst all stressful life events exert demand on an individual's biological or psychological capabilities, they can be sub-divided into those that are 'chronic', such as enduring environmental disadvantage or recurring stressful life events, and those that are 'acute', involving discrete change in an individual's current environmental conditions. Acute stressful life events can be further divided into single specific events which exert independent effects (e.g. life transitions), or the cumulative effects of numerous events occurring over a specific period (e.g. major life events and daily hassles). Major life events refer to significant events impacting an individual's functioning (e.g. death of a loved one), whilst daily hassles refer to distressing everyday events (e.g. peer conflict).

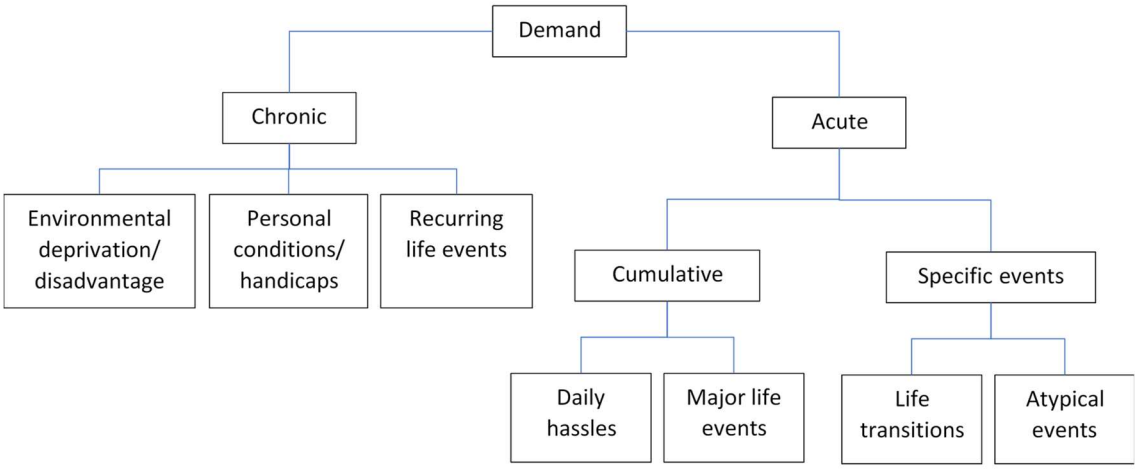


Figure 1. Conceptualisation of stressful life events (taken from Compas, 1987)

Two main models for understanding the impact of stressful life events on individuals' physical and psychological well-being have also been proposed (Cohen, Kessler et al. 1997). The transactional or "subjective" approach views stress as the result of individuals' cognitive appraisal of the type and severity of distress caused by specific environmental events or conditions (Lazarus 1990). It emphasises the individuals' subjective interpretation of the stressful life events rather than their experience *per se*. Alternatively, the "objective" approach is concerned with exposure to environmental events that represent quantifiable change in characteristics of the individuals' environment. More specifically, Grant and colleagues (2003) defined stressful life events as "*environmental events or chronic conditions that objectively threaten the physical or psychological health or well-being of individuals of a particular age and in a particular society"*. This approach highlights the importance of identifying external environmental change that is distinct from any potential moderating and mediating cognitive processes, such as individual differences in judgement of how threatening the event is.

Stressful life events and developmental psychopathology

Regardless of the conceptualisation, many theoretical models of developmental psychopathology emphasise experiences of stressful life events as important risk factors for the development and maintenance of psychological problems in childhood and adolescence. For example, developmental models of depression highlight the importance of early childhood experiences (e.g. of parental neglect, abuse or loss, disturbed family environments) in subsequent development of depression (Hammen and Rudolph 1996).

Childhood and adolescence are especially important developmental periods for experiences of life stress. The dynamic nature of child development means that young people are likely to experience changes and transitions across multiple life domains, increasing the likelihood of exposure to stressful life events. Adolescence represents a major developmental transition period in which life stress is particularly common (Compas 1987, Kim, Conger et al. 2003). It is a time when adolescents gain increased independence from parents and place more emphasis on peer relationships as well as experiencing greater academic pressures. The increase in stressful life events in adolescence is suggested to contribute to the large increase in the incidence of depression also observed at this time (Ge, Lorenz et al. 1994, Ge, Conger et al. 2001). Additionally, due to developmental changes in an individual's ability to cope with or adapt to stressful life events (e.g. with transitions in support networks), experiences in childhood may have a more pervasive impact on psychological health than events experienced in adulthood. (Rutter 1981, Grant, Compas et al. 2003). As a result,

understanding the impact of stressful life events on the development of psychological disorders in children and adolescents is vitally important for informing interventions aimed at reducing young people's exposure to stressors or at enhancing young people's abilities to manage stressful experiences.

In adult populations, stressful life events have been shown to be associated with a range of negative outcomes in both physical (Tosevski and Milovancevic 2006) and psychological indices (Dohrenwend 2006). However, despite the potential for stressful life events to contribute to the development of psychopathology in childhood and adolescence, supporting research in young people is generally lacking. A review of the stress-symptoms literature in young people is beyond the scope of this study, however, several key themes have been identified. There is robust evidence for prospective associations between severe traumatic experiences (e.g. major accident/injury, childhood abuse) and subsequent onset of post-traumatic stress disorder, as well as more general psychopathology, in youth (Copeland, Keeler et al. 2007, Hovens, Wiersma et al. 2010, Trickey, Siddaway et al. 2012). Using measures of major life events and daily hassles, there is also consistent evidence for prospective associations between the cumulative effects of multiple stressful life experiences and elevated internalising symptoms, particularly depression and anxiety, in young people (Grant, Compas et al. 2004). However, although limited, research also suggests that daily hassles may be more associated with the subsequent development of depression than major life events (Kanner, Coyne et al. 1981, Compas, Davis et al. 1987, Seiffge-Krenke 2000). Few studies in the young have compared both major and daily hassles events within a single study (Compas, Davis et al. 1987).

Additional published research demonstrates varying forms of specificity in associations between certain types of life events and internalising symptoms. Studies in both adults and children have suggested that threat events (e.g. physical jeopardy) are more linked with anxiety, and that loss events (e.g. loss of attachment figure) are specifically associated with depression (Finlay-Jones and Brown 1981, Eley and Stevenson 2000).Within studies focussing on depression in young people, there is evidence that dependent events, those that are in some way related to characteristics of the individual, and particularly interpersonal dependent events (e.g. peer conflict), are more related to depression symptoms than independent events (e.g. parental divorce) (Williamson, Birmaher et al. 1995, Kercher, Rapee et al. 2009, Hamilton, Stange et al. 2013). In adults, studies have identified a possible protective role of positive life events (e.g. satisfying social interactions,

pleasant activities) as a buffer against psychological problems (Cohen, McGowan et al. 1984, Dixon and Reid 2000).

Finally, moderating and mediating factors that affect the relationship between stressful life events and psychopathology have been proposed (Figure 2). For example, the stress reactivity hypothesis of depression which proposed that personal characteristics (e.g. gender, cognitive style) moderate reactivity to stress, was then proposed to subsequently predict depression severity in youth (Compas, Connor-Smith et al. 2004, Hankin, Mermelstein et al. 2007). Studies exploring potential mediating factors have looked at various psychological processes, including coping skills (Chance 1982), emotion regulation (Stikkelbroek, Bodden et al. 2016) and parental factors (Platt, Williams et al. 2016). Research has also suggested the possibility of reciprocal relationships between stressful life events and psychopathology. For example, using the stress generation model of depression, vulnerabilities for depression (prior depressive episode, depressive cognitive style) were shown to be associated with subsequent experiences of stressful life events in youth, which in turn predict increases in depression symptoms (Hamilton, Stange et al. 2013).

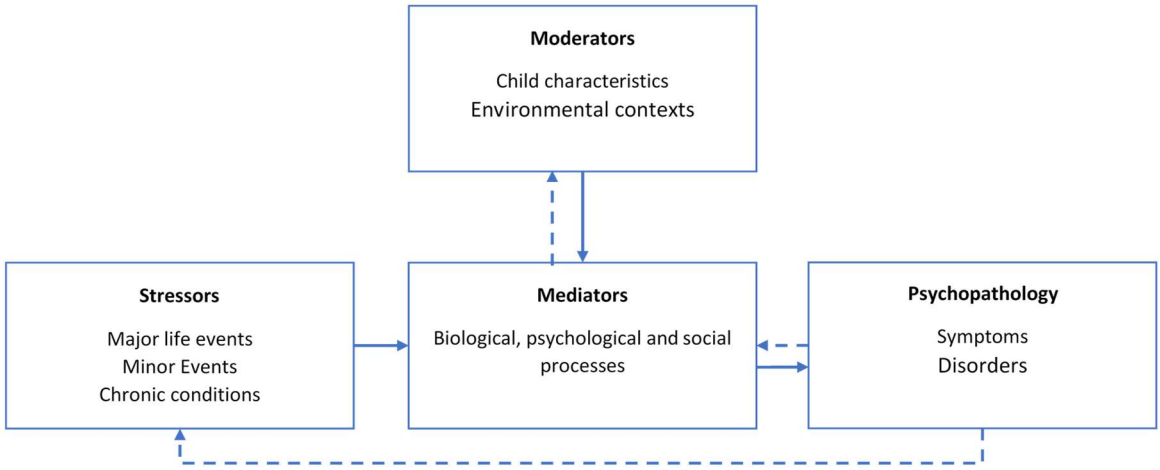


Figure 2. Conceptual model of the relationship between stressful life events and psychopathology in youth (taken from Grant et al., 2003)

The studies identified suggest there are complex and dynamic associations between stressful life events and psychological problems across development, and that a variety of personal variables (e.g. age, gender, cognitive style) and contextual factors (e.g. type of life event) alter the association between stressful life experiences and individual wellbeing.

Measurement of stressful life events in youth

Given the complex interplay between young people's experiences of stressful life events and psychopathology, the accurate measurement of stressful life events is essential in broadening our understanding of the dynamic influences of such experiences on childhood outcomes. However, differences in the way stressful life events have been conceptualised and measured across studies has made it difficult to compare research findings and to confidently draw conclusions about the role of stressful life events in the development of psychopathology in youth. Studies vary in assessment method, who is selected as informant, what response format is used, and which type of stressful life event is captured.

Assessment method

Some researchers have developed interview-based measures for young people, designed to provide objective assessment of the contextual threat associated with stressful events (e.g. Rudolph and Hammen 1999, Duggal, Malkoff-Schwartz et al. 2000). These typically use individual responses to probe questions to generate a list of encountered life events together with the contextual factors surrounding them. The level of "stress" associated with each event is evaluated by multiple raters and then summed to give total stress scores. Whilst interview-based measures can provide rich contextual accounts of stressful life events, they are time-consuming to administer and code. Consequently, the most widely used method for assessing stressful life events affecting children and adolescents are checklists covering a range of typically occurring stressful life events.

Informant

Measurement of the incidence and effects of stressful life events in children and adolescents tends to use self-reports or reports from other close adults, such as parents and teachers. Few studies have examined informant effects (Bailey and Garralda 1990, Rende and Plomin 1991, Johnston, Steele et al. 2003). These few studies show that children (typically in middle childhood) generally report higher frequencies of negative events than their parents but significantly lower distress ratings in relation to these events. Furthermore, whilst parents may provide better information on stressors relating to the family context (e.g. parental health, wealth and relationships), children are likely to report stressful experiences in the context of peer relationships more accurately (Rende and Plomin 1991). This is particularly important for measuring stressful life events in adolescents, who spend relatively more time away from their parents than younger children.

Response format

The measures used also vary in their response format. Some ask respondents to indicate whether they have experienced specific events within the given time frame (occurrence) whilst others ask how often (frequency) or how severe (impact) the event was. For example, the Children's Hassles and Uplifts Scales (Kanner, Feldman et al. 1987) asks young people to indicate which events they had experienced (occurrence) and then to rate how bad or good they felt (impact) about the endorsed items. Both frequency and impact ratings involve a degree of subjective appraisal of the stressful event, in line with transactional conceptualisations of stressful life events (Cohen, Kessler et al. 1997) whilst occurrence ratings are in line with objective conceptualisations of stressful life events (Grant, Compas et al. 2003) and remove the possible confound with mediating cognitive appraisal variables.

Types of stressful life events

There is also variation in the type of stressful life events captured. Some measures capture major life events (e.g. parental separation, illness or death) typically occurring over the previous 6-12 months (e.g. Coddington 1972, Williamson, Birmaher et al. 2003) whilst others aim to measure daily hassles (e.g. arguments with friends or family) over brief intervals of weeks or months (e.g. Kanner, Feldman et al. 1987, Shahar, Henrich et al. 2003). There is also variation in the life domains that are captured with different subscales identified; for example, reflecting different settings (e.g. home, school) or contexts (e.g. peer, family, academic) or broader dependence-independence or interpersonal-non-interpersonal domains. It is also common for researchers to have designed bespoke measures of life events for use with specific samples (e.g. scales developed specifically for rural or urban populations (Miller and Townsend 2005)).

Previous reviews of life event measures in young people

The specific measures selected for individual studies vary widely depending on the research hypotheses, target populations and researcher preference. Measures are often altered by adapting or excluding certain items or adding additional items to suit specific study aims. The psychometric properties of these measures, and indeed the original measures from which they were derived, have been rarely examined in detail.

Grant and colleagues (2004) identified over 500 studies examining stressful life event measures and symptoms of psychopathology in children and adolescents. Of these, 45% reported developing measures themselves with the remaining 55% selecting one of over 50

available measures. The authors noted that the method for developing the measures and psychometric data for these measures were rarely reported.

A taskforce of the American Psychological Association examined measures of both coping and stress in young people by asking members of their survey list to identify measures that they commonly used (Blount, Simons et al. 2007). Measures were included if they were endorsed by five or more of the society's members, resulting in the inclusion of only three measures of stress (but 12 measures of coping). Measures were classified by six raters as 'well-established', 'approaching well-established' or 'promising' depending on number of criteria, including the number of peer-reviewed articles and evidence of assessing psychometric properties. Only one self-report measure of children's stress, the Children's Hassles and Uplifts Scale (Kanner, Coyne et al. 1981), was considered 'well established' whilst two further measures (Coddington Life Event Scales, Coddington 1972; Questionnaire on Resources and Stress, Holyroyd, 1974) required further investigation of their psychometric properties.

These two previous reviews highlight the range of measures of life events in young people available to researchers, but they also indicate a lack of methodological rigour with which the measures have been developed and their psychometric properties have been assessed. Whilst Grant and colleagues' (2004) review highlights the limited assessment of psychometric properties, this was not the main focus of the review and so, they did not explicitly examine the psychometrics of individual measures necessary for informing researchers for selection of scales for their studies. Additionally, Blount and colleagues' (2007) reliance on endorsement by society members could potentially bias the selected measures towards those favoured by certain researchers. Newer measures may not have been endorsed because they were less well known rather than less valid or useful. Finally, both reviews were conducted over 15 years ago. There has been considerable development in the field since their publication and several new measures have been developed. A comprehensive systematic review of existing measures of stressful life events in young people that considers psychometric properties has not been conducted to date.

The broad range of available measures and a lack of understanding of their psychometric properties makes it difficult for researchers to select the most appropriate measure for their own research. Understanding the psychometric properties of the different measures of stressful life events used for young people is vitally important for determining whether the

results are due to meaningful differences within or across stressful life event measures (e.g. due to major life events or daily hassles) or are artefacts of measurement.

Current systematic review

A systematic literature review with specific focus on the methodology used to develop measures of stressful life events and their psychometric properties was conducted to provide a comprehensive appraisal of how the effects of stressful experiences in young people are assessed. It is hoped this review will facilitate researchers' selection of the most appropriate measure of stress for their study aims and reduce inconsistency in how measures are developed or adapted for research, making it easier to compare findings and draw conclusions across studies. Ultimately, this will allow greater understanding of the interplay between stress and mental health outcomes in young people which can potentially inform intervention and prevention strategies.

Methods

The current review captured studies reporting on psychometric properties of stressful life event measures used with people 18 years or younger. This review followed the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines (Moher, Shamseer et al. 2015); an evidence-based set of minimum reporting standards for systematic reviews and meta-analyses, including criteria for reporting methods (e.g. search terms, selection criteria, risk of bias analysis) and results (e.g. study selection numbers, study characteristics, summary data).

Searching and inclusion/exclusion criteria

Three computerised databases (PsychInfo, Medline, Web of Science) were searched for relevant articles using Ovid SP. The search included studies published up to 2nd June 2016. Three categories of index terms were used relating to "stressful life events", "psychometrics" and "psychopathology". Wildcards and truncation were used to maximise search efficiency. Searches were restricted to children and adolescents (See Appendix 1 for PsychInfo search strategy). No publication date or language restrictions were set. Manual searches of reference lists of relevant articles were conducted to ensure all applicable studies were included. Studies that only included participants over 18 years were excluded. Only measures of discrete stress events, and not symptoms or the impact of stress itself, were included. Where multiple studies pertaining to the same measures were obtained, only the initial development study was included. This was a necessary limitation to ensure a

feasible number for systematic review and synthesis. Review papers and studies that did not describe the initial development of the measure were excluded but attempts were made to source original manuscripts. Studies were excluded if full-text could not be retrieved or was not translated into English.

All titles and abstracts were screened to identify eligible studies. Full-text versions of eligible studies were downloaded into EndNote X7 (Thomson Reuters 2016). Studies were included if they reported any of the common psychometric properties, including content validity, internal consistency, reliability, convergent validity, criterion validity or cross-cultural validity. Each psychometric property is described in more detail in Table 1 (Mokkink, Terwee et al. 2010).

Methodological Quality Assessment

All eligible studies were critically appraised using the Consensus-based Standards for the selection of health Measurement Instruments (COSMIN; Mokkink, Prinsen et al. 2016). The COSMIN taxonomy outlines criteria for evaluating the methodological quality of commonly assessed psychometric properties in health-related patient outcome measures; namely, content validity, internal consistency, test-retest reliability, structural validity, criterion validity, cross-sectional validity and hypothesis testing. Each psychometric property domain is operationalised and rated as 'poor', 'fair', 'good' or 'excellent' quality based on the presence of certain criteria (see Table 1 for details).

Each paper was independently rated by two reviewers (the lead author, HB and their supervisor). If the study did not report sufficient detail to allow a rating, then it was rated as not assessed. Agreement was reached for 92.3% of all ratings. Discrepancies were discussed until consensus was reached. Overall methodological quality ratings for each psychometric property for each paper were given using the lowest rating of any criteria within that domain, i.e. "worst score counts" (Terwee, Bot et al. 2007). Since one of the aims of the current systematic review was to assess the quality of methods used to develop stressful life event measures, all measures were included in the final review, regardless of assigned quality ratings.

Table 1. Definitions of psychometric properties (adapted from Mokkink et al., 2010)

Psychometric Property	Definition	COSMIN Criteria for "excellent" ratings
Content validity	The degree to which the content of the instrument is an adequate reflection of the construct to be measured	1. All items refer to the relevant aspects of the construct to be measured and area assessed as relevant to the study population e.g. using a focus group (sample size >10) 2. All items relevant for the application of the measure 3. All items together comprehensively reflect the construct to be measured
Internal Consistency	The degree of interrelatedness between items	1. Scale consists of effect indicators i.e. based on a reflective model 2. Adequate sample size (> 100) 3. Unidimensionality checked using factor analysis 4. Internal consistency statistic calculated for each (sub)scale 5. Appropriate statistical methods; Cronbach's Alpha for continuous scores
Test-retest reliability	The extent to which scores on the same version of questionnaires for people who have not changed are the same for the repeated measurement over time	1. Adequate sample size (> 100) 2. At least to measurements available 3. Independent administrations 4. Participants stable between administrations and similar test conditions 5. Appropriate time interval 6. Appropriate statistical methods; intraclass correlations coefficient for continuous scores
Structural validity	The degree to which scores of an instrument are an adequate reflection of the dimensionality of the construct to be measured	1. Scale consists of effect indicators i.e. based on a reflective model 2. Adequate sample size (> 100 or 5-7 x no. of items but <100) 3. Exploratory or confirmatory factor analysis performed appropriately in view of existing information
Cross-cultural validity	The degree to which the performance of items on a translated or culturally adapted instrument are an adequate reflection of the performance of the items of the original version of the measure	1. Adequate sample size (7* no of items and ≥ 100) 2. Both source and target language described 3. Independent translator with adequate expertise 4. Forward and backward translation and describes how differences were resolved. 5. Pre-tested in target population
Criterion Validity	The correlation of the instrument with a "gold standard" criterion administered at the same time	1. Adequate sample size (>100) 2. "Gold standard" criterion used 3. Appropriate statistical methods: Correlations used for continuous scores
Hypothesis testing	The degree to which scores of the instrument relate to scores on other measures to which it should be related	1. Adequate sample size (>100) 2. Multiple hypotheses formulated a priori including expected direction and magnitude of effect 3. Adequate measurement properties of the comparator instrument(s) in a population like the study population 4. Appropriate statistical methods for the hypotheses to be tested

Data extraction

Quantitative data for each psychometric property was extracted from each eligible study. The following established criteria were used to assess 'acceptability' for each psychometric property for each measure (Hu and Bentler 1999, Terwee, Bot et al. 2007, Kersten, Czuba et al. 2016): Internal Consistency: Cronbach's alpha $\geq$ .70; Test-retest Reliability: Correlation coefficients $\geq$ .80; Structural Validity: Confirmatory Factor Analysis (CFA): RMSEA < .06 good fit, < .08 acceptable fit; CFI, GFI etc. > .9 good fit .8 - .9 acceptable fit; Criterion Validity: $r \geq$ .70. Narrative summaries with reference to best-practice guidelines were provided for psychometric properties without quantitative criteria; namely content validity, hypothesis testing and cross-cultural validity.

Results

Summary of included studies

The systematic search resulted in a total of 21 papers meeting inclusion criteria (Figure 3). Of these, 20 studies described distinct self-report life event checklists and one detailed an interview measure with parallel self- and parent-reports (Table 2). Age ranges varied across studies with some taking a narrow focus on just one year group (e.g. Children's Hassles and Uplifts Scale; Kanner, Feldman et al. 1987) and others covering childhood and adolescent years more broadly (e.g. Child and Adolescent Survey of Experiences; Allen, Rapee et al. 2012). The youngest age groups were children aged 8 – 11 years (Children's Own Perceptions and Experiences of Stressors; Colton 1985) and 7 – 14 years (Daily Life Stressors Scale; Kearney, Drabman et al. 1993). The majority of studies used samples of unselected, healthy children whilst three studies compared 'disadvantaged' (Kearney, Drabman et al. 1993, Miller, Webster et al. 2002) or 'psychiatrically ill' children (Williamson, Birmaher et al. 2003) and controls.

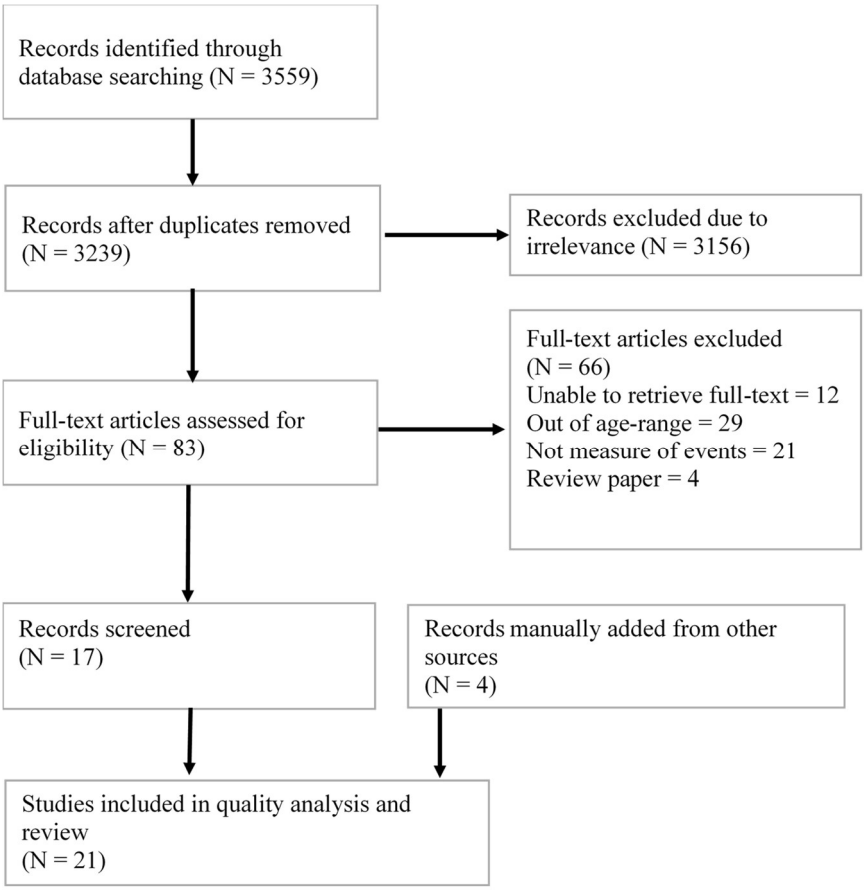


Figure 3. PRISMA flow chart of search results

The checklist measures varied in the time frame for life events captured and in the response scale used. Some focussed on major life events (e.g. Social Readjustment Scale; Coddington 1972) whilst others focused on daily hassles (e.g. Adolescent Hassles Inventory; Bobo, Gilchrist et al. 1986). Some did not attempt to distinguish between them. Measures of major life events tended to refer to longer time frames (typically 6 – 12 months), whilst daily hassle measures typically referred to periods of 1- 4 weeks. A few measures did not explicitly state a time frame in which events must have occurred (e.g. Bagley 1992, Nilsson, Gustafsson et al. 2010). The number of items included in measures varied widely from brief measures of only 9 events (Urban Hassles Scale; Miller, Webster et al. 2002) up to those with 320 events (Unpleasant Events Schedule; Lewinsohn and Talkington 1979). Studies also varied in the response format used. The majority asked whether each event had *occurred* within the given time frame (n = 11) whilst others used *frequency* (n = 6) or *impact* (n = 4) ratings using various Likert scales.

Table 2. Summary of included measures (in order of publication date)

Measure (Authors)	Informant / Population	Context/ Time frame	No. of items	Response Format	Factors/Subscales
Social Readjustment Rating Questionnaire (Coddington, 1972)	Self-/Parent-report, age not given, Healthy children	Major life events, 12 months	35	Occurrence	--
Unpleasant Events Schedule (Lewinsohn & Talkington, 1978)	Self-report, 15-85+, Depressed and controls	Typical events, 30 days	320	Frequency: 3-point scale: 1(not happened) to 3 (often - 7+ times) Aversiveness: 3-point scale: 1(not unpleasant) to 3 (very unpleasant)	Health and well-being Achievement-academic-job Domestic, Day-to-day inconveniences Sexual, Marital, Friendship Legal Material-Financial Social Exits
Children's Own Perceptions and Experiences of Stressors (Colton, 1985)	Self-report, 8-11 years, Healthy children	Hassles and major events, Lifetime	60	Occurrence Impact: 5-point scale: 1 (not upsetting) to 5 (extremely upsetting)	Isolation Major life events Cognitive overload Family disruption Financial Step-families Serious school problems
Adolescent Hassles Inventory (Bobo et al., 1986)	Self-report, 11-12 years, Healthy children	Daily hassles, 1 week	50	Occurrence Impact: 3-point scale: 1 (small/minor) to 3 (large/major)	School Peers Future Drugs Parents Work Bother Money
Adolescent Perceived Events Scale (Compas et al., 1987)	Self-report, 12-20 years, Healthy children	Negative and Positive events, 3 months	157 (+36 for young/mid adolescents, +15 for late adolescence)	Impact: 9-point scale: -4 (extremely bad) to +4 (extremely good)	Positive/Negative Life events/hassles
Children's Hassles and Uplifts Scale (Kanner et al., 1987)	Self-report, 11-12 years	Daily hassles, 1 month	25 Hassles 25 Uplifts	Occurrence Frequency: 3-point scale: 0	Hassles Uplifts

				(didn't feel bad) to 2(felt very bad)	
Life Events and Coping Inventory (Dise-Lewis, 1988)	Self-report, 11-14 years, Healthy children	Daily hassles and major life events, 12 months	125	Occurrence Impact: 9-point scale: 0 (not stressful) to 9 (extremely stressful)	--
Adolescent Stress Questionnaire (Bagley, 1992)	Self-report, 14 - 16 years, Healthy children	Typical stress events, Lifetime	68	Impact: 4-point scale: 1(no stress) to 4 (extreme stress)	Relationship problems Abuse at home (drugs), Scholastic problems, Health/Personal concerns, Peer Pressure, Family problems, Loneliness/Isolation
Daily Life Stressors Scale (Kearney et al. 1993)	Self-report, 7-14 years, Children's/foster home and controls	Daily hassles, 1 week	30	Occurrence Impact: 4-point scale: 0 (not at all stressful) to 3(a lot)	--
Chinese Adolescent Life Event Scale (Cheng, 1997	Self-report, 12-18 years, Healthy children	6 months	44	Occurrence Desirability: 7-point scale: 1(extremely undesirable) to 7 (extremely desirable) Impact: 7-point scale: 1(extremely small impact) to 7 (big impact)	--
Urban Hassles Scale (Miller et al., 2002)	Self-report, 14-19, Disadvantaged children	Daily hassles	9	Frequency: 3-point scale: 0(none) to 2 (a lot)	--
Brief Adolescent Life Event Scale (Shahar et al., 2003)	Self-report, 12-15 years, Healthy children	Daily positive and negative events, 4 weeks	36	Frequency: 4-point scale: 0 (never) to 3 (a lot)	Family life Friendship Peer relations Extracurricular activities Performance Health & Physical appearance (negative and positive scales)
Stressful Life Event Schedule (Williamson et al., 2003)	Self-/Parent-interview, 11-15 years, psychiatrically ill and controls	12 months	-- (interview)	Occurrence + follow-up questions to contextualise	
Urban Hassles Scale (Revised) (Miller et al., 2005)	Self-report, 10-20, Healthy children	Daily Hassles, 2 weeks	32	Frequency: 4-point scale: 0 (never) to 3 (very often)	Environmental conditions Interpersonal

					interactions/Surveillance Safety concerns Anticipatory victimisation
Adolescent Stress Questionnaire (modified) (Byrne et al., 2007)	Self-report, 13-18 years, Healthy children	Typical negative events, 12 months	58	Impact: 5-point scale: 1 (not at all stressful) to 5 (very stressful)	Stress of home life Stress of school performance Stress of school attendance Romantic relationships Peer pressure Teacher Interaction Future Uncertainty School-Leisure Conflict Financial Pressure Adult responsibilities
Hassles Measure for Indian Adolescents (Mehrotra & Kumari, 2009)	Self-report, 15-20 years, Healthy children	Hassles, 6 months	28	Frequency: 5-point scale: 1 (never) to 5 (almost always)	--
Adolescent Stress Questionnaire - Norwegian (Moksnes et al., 2010)	Self-report, 13-18 years, Healthy children	Typical negative events, 12 months	58	Impact: 5-point scale: 1 (not at all stressful) to 5 (very stressful)	As with Byrne et al., (2007)
Linkoping Youth Life Experience Scale (Nilsson et al., 2010)	Self-report, 15-19 years, Healthy children	Typical and major negative events, Lifetime	41 (23 main items with secondary items)	Occurrence: Checklist	Non-interpersonal Interpersonal Long-standing adverse childhood circumstances
Brief Daily Hassles Scale (Wright et al., 2010)	Self-report, M_{age}=15.35, Healthy children	Daily hassles, 4 weeks	17	Frequency: 5-point scale: 0 (never) to 5 (daily)	Hassles from parents Hassles from peers/others
Child and Adolescent Survey of Experiences (Allen et al., 2012)	Self-/Parent-report, 7- 16 years, Healthy children	Positive and negative events, past 12 months	38	Occurrence: Checklist Impact: 6-point scale: 1 (really good) to 6 (really bad)	--
Hispanic Stress Inventory - Adolescent (Cervantes et al., 2012)	Self-report, 10-20 years, Healthy children	Typical events, Lifetime	160	Occurrence Impact: 5 - point scale: 1(not worried) to 5 (extremely worried)	Family economic stress Culture and education stress Acculturation stress Immigration stress Discrimination stress Family immigration stress Community/gang-related stress Drug-related stress

Finally, the number and nature of subscales defined in studies varied. Eight studies referred to unidimensional measures of stressful events. The rest defined between two and 10 subscales, capturing different aspects of stressful experiences; the majority depicting different life domains (e.g. school, peers, family) but a few captured the degree of dependence on the individual or the interpersonal nature of the events (e.g. Linkoping Youth Life Experience Scale; Nilsson, Gustafsson et al. 2010).

Psychometric properties of the life event measures

Summary

The identified studies varied in the psychometric properties assessed (Table 3) with an average of three psychometric properties examined per study. Most papers (n = 19) described, to a lesser or greater extent, how the measure was developed (content validity). Of the 21 included studies, 12 examined internal consistency and 10 reported test-retest reliability. Structural validity was examined in 14 of the included papers. Although most papers (n = 17) examined associations between the target life event measure and some hypothesised variable(s), very few (n = 5) compared the target life event measure and a more-established life event measure (criterion validity). Only one study examined cross-cultural validity. According to ratings on the COSMIN checklist, more recent studies had a greater proportion of 'good' or 'excellent' methodological qualities ratings than early papers. Findings for each psychometric property are discussed in detail below.

Table 3. Methodological quality ratings (COSMIN) for each psychometric property examined by paper (date order)

Not examined	Poor	Fair	Good	Excellent

Measure	Content Validity	Internal consistency	Test-retest reliability	Structural Validity	Hypothesis testing	Cross-cultural validity	Criterion Validity
Social Readjustment Rating Questionnaire (Coddington, 1972)	Poor						
Unpleasant Events Schedule (Lewinsohn & Talkington, 1978)	Good	Fair	Fair	Poor	Excellent		
Children's Own Perceptions and Experiences of Stressors (Colton, 1985)	Excellent			Poor			
Adolescent Hassles Inventory (Bobo et al., 1986)	Fair	Fair	Fair	Poor	Poor		
Adolescent Perceived Events Scale (Compas et al., 1987)	Excellent		Fair				Fair
Children's Hassles and Uplifts Scale (Kanner et al., 1987)	Fair	Fair		Poor	Fair		
Life Events and Coping Inventory (Dise-Lewis, 1988)	Good		Fair	Poor	Poor		
Adolescent Stress Questionnaire (Bagley, 1992)	Poor	Poor		Fair	Fair		
Daily Life Stressors Scale (Kearney et al. 1993)	Poor		Fair		Poor		
Chinese Adolescent Life Event Scale (Cheng, 1997)	Good		Fair				Fair
Urban Hassles Scale (Miller et al., 2002)	Fair	Fair		Fair			
Brief Adolescent Life Event Scale (Shahar et al., 2003)	Poor	Fair		Fair	Fair		
Stressful Life Event Schedule (interview) (Williamson et al., 2003)	Poor		Fair		Fair		Fair
Urban Hassles Scale (Revised) (Miller et al., 2005)	Good	Excellent		Excellent	Excellent		
Adolescent Stress Questionnaire (modified) (Byrne et al., 2007)	Good	Good	Fair	Good	Fair		
Hassles Measure for Indian Adolescents (Mehrotra & Kumari, 2009)	Good	Poor			Fair		
Adolescent Stress Questionnaire – Norwegian (Moknes et al., 2010)		Good		Good	Excellent	Fair	
Linkoping Youth Life Experience Scale (Nilsson et al., 2010)			Fair		Fair		
Brief Daily Hassles Scale (Wright et al., 2010)	Good	Fair		Poor	Fair		
Child and Adolescent Survey of Experiences (Allen et al., 2012)	Excellent		Fair	Poor	Fair		Fair
Hispanic Stress Inventory – Adolescent (Cervantes et al., 2012)	Good	Good		Good	Excellent		

Content Validity

Of the 21 included studies, 19 described scale development in sufficient detail to assess their content validity. Methodological quality ratings according to COSMIN criteria ranged from 'poor' to 'excellent' (Table 3, Column 2). The highest rating is given to measures using "bottom-up" methods of rating scale development where young people in the target age range or experts outside of the immediate research team are consulted at the initial inception of the measure and then subsequently consulted on the relevance of the finalised items.

Only 3 of the included studies (shown in blue in column 2, Table 3) involved adolescents in both item generation and assessing item relevance. For example, Colton (1985) used top-down and bottom-up item generation by conducting a literature review of existing child and adult measures and then holding a focus group with children to identify further relevant items before piloting the measure with an independent group of children to assess typical endorsement of included items.

Studies rated as 'poor' (n = 5, shown in red in Table 3, Column 2) tended to use "top-down" approaches to scale development by selecting and adapting items from existing measures or based on the researcher knowledge. For example, Bagley (1992) added additional items to an existing measure to reflect his specific interest in child abuse and removed items deemed to be specific to ethnic minorities due to the rural context of their research. Kearney and colleagues (1993) derived items from the author's clinical experiences of assessing young people and anecdotal self-reports of young people in the clinic.

Internal Consistency

Only 12 of the 21 included papers reported the internal consistency of the proposed measure. Methodological quality ratings ranged from 'poor' to 'excellent' (Table 3, Column 3). Under COSMIN criteria, 'poor' ratings were given if the internal consistency of proposed subscales was not reported or sample size was inadequate (n = 2; shown in red, Colum 3, Table 3). All nine studies reporting Cronbach's alpha for the total scale demonstrated acceptable internal consistency (Table 4). Of eight studies proposing life event subscales, seven reported their internal consistency. Cronbach's alpha values were generally lower for the subscales of life event measures, ranging from .60 - .92. Internal consistency less than .70 indicates a lack of correlation between items in the scale and questions the justification of summarising constituent items within a single subscale. Lower internal consistency for subscales may also be, in part, due to few items included within the subscales (some having as few as three items).

Test-retest reliability

Ten of the included studies examined test-retest reliability. The retest interval varied across the included studies, ranging between 5 days (Shahar, Henrich et al. 2003) and 3 months (Lewinsohn and Talkington 1979). Methodological quality ratings were all 'fair' (Column 4, Table 3). According to the COSMIN checklist, quality ratings were lower because intra-class correlations (ICCs) were not used. ICC is considered the most suitable statistic for assessing test-retest reliability of continuous measures because it takes into account systematic variance differences (Terwee, Bot et al. 2007). However, only one study (Brief Adolescent Life Events Scale; Shahar, Henrich et al. 2003) used ICCs. The rest either used Pearson's correlation coefficient for total scale/subscale scores (n = 7) or examined item-level agreement using Cohen's Kappa coefficient (n = 3) as recommended for ordinal level data (Table 4). Regardless of statistical method used, Terwee and colleagues (2007) recommend a minimum reliability coefficient of .70. All 10 studies demonstrated acceptable test-retest reliability for total scale scores at their selected retest interval. However, test-retest coefficients were often lower for constituent subscales with some below the .70 threshold. For example, Bobo and colleague's (Adolescent Hassles Inventory; 1986) estimates ranged from .49 for their 'money-related' events subscale and .81 for 'peer' events.

Some studies also examined differences in test-retest reliability as a function of certain demographic characteristics. For example, Williamson and colleagues (Stressful Life Event Schedule; 2003) found similar reliability coefficients in subgroups of children compared to adolescents but greater stability in controls compared to young people with psychiatric illness.

Table 4. Internal consistency and test-retest reliability of stress measures in young people

			Test-retest reliability		
	Variables/ Subscales	Internal consistency	Statistic used	Interval	Value
Social Readjustment Rating Questionnaire (Coddington, 1972)	Frequency of events	--	--	--	--
Unpleasant Events Schedule (Lewinsohn & Talkington, 1978)		.96	Pearson's *r*	1 month*	.67
	Frequency/Impact of events			2months	.68
	8 subscales:	.76		3 months	.64
	- Health and wellbeing	.81			
	- Achievement-Academic-Job	.74			
	- Domestic, Day-to-day inconveniences	.67			
	- Sexual, Marital, Friendships	.70			*reliability coefficient for total score (subscale stabilities at each interval are also reported in paper)
	- Legal	.70			
	- Material-Financial	.70			
	- Social Exits	.60			
	- Most Discriminating Items	.82			
Children's Own Perceptions and Experiences of Stressors (Colton, 1985)	Frequency/Impact of events 7 subscales: - Isolation - Major life events - Cognitive Overload - Family disruption - Financial - Step-families - Serious school problems	--	--	--	--
Adolescent Hassles Inventory (Bobo et al., 1986)	Frequency/Impact of events		Pearson's *r*	1 week	
	8 subscales				
	- School	.84		Subsample (N = 116)	.62
	- Peers	.85			.81
	- Future	.75			.73
	- Drugs	.78			.70
	- Parents	.74			.71
	- Work	.73			.60
	- Bother	.67			.73
	- Money	--			.49
Adolescent Perceived Events Scale (Compas et al., 1987)	Negative, Positive Events (Frequency, Impact)	--	Pearson's *r*	2 weeks	Young (12-14): .85 (freq.) / .86 (neg weighted) / .78 (pos weighted) Mid (15-17): .84 (freq) / .89 (neg weighted) / .81 (pos weighted) Late (18-20): .77 (freq) / .74 (neg weighted) / .84 (pos weighted)
Children's Hassles and Uplifts Scale (Kanner et al., 1987)	Frequency/Impact of events 2 subscales:		--	--	--
	- Hassles	.87			
	- Uplifts	.87			
Life Events and Coping Inventory (Dise-Lewis, 1988)	Frequency/Impact of events	--	Pearson's *r*	11 weeks	.97 (Individual items: .07 - .56, median = .25)

Adolescent Stress Questionnaire (Bagley, 1992)	Impact 7 subscales: Relationship problems, Abuse at home: drug/alcohol, Scholastic problems, Health/Personal concerns, Peer Pressure, Family problems, Loneliness/Isolation	.83 (total) (subscale alphas not reported)	--	--	--
Daily Life Stressors Scale (Kearney et al., 1993)	Frequency/Impact of events	--	Pearson's *r*	1 week Subsample of 2nd grade only (n = 97)	.74 (Individual items: .16 - .71)
Chinese Adolescent Life Event Scale (Cheng, 1997	Frequency of events	--	Kappa	2 weeks	0.86
Urban Hassles Scale (Miller et al., 2002)	Frequency of events	.08	--	--	--
Brief Adolescent Life Event Scale (Shahar et al., 2003)	Frequency Negative, Positive Events 6 subscales - Family Life - Friendships - Peer relations - Extra-curricular activities - Performance - Health/Physical appearance	.87 (neg) / .88 (pos) .60 - .80 (subscales)	--	--	--
Stressful Life Event Schedule (Williamson et al., 2003)	Occurrence	--	ICC	5-15 days	.93 (total sample) 91 (child) / .94 (adol.) .83 (psych illness) / .93 (control)
Urban Hassles Scale (Revised) (Miller et al., 2005)	Frequency 4 subscales: - Environmental conditions - Interpersonal Interactions - Safety concerns - Anticipatory victimisation	.85 .77 .75 .75 .69	--	--	--
Adolescent Stress Questionnaire (modified) (Byrne et al., 2007)	Impact 10 subscales: - home life stress - school performance stress - school attendance stress - Romantic relationships - Peer pressure - Teacher Interaction - Future Uncertainty - School-Leisure Conflict - Financial Pressure - Adult responsibilities	-- .92 .88 .69 .86 .88 .87 .82 .86 .83 .62	Pearson's *r*	1 week	 .88 .81 .86 .84 .86 .80 .79 .81 .68 .77
Hassles Measure for Indian Adolescents (Mehrotra & Kumari, 2009)	Frequency of events	.78	--	--	--

Adolescent Stress Questionnaire - Norwegian (Moknes et al., 2010, 2011)	10 subscales (as in Byrne et al., 2007)	.70 - .89	--	--	--
Linkoping Youth Life Experience Scale (Nilsson et al., 2010)	Frequency of events - Interpersonal - Non-interpersonal - Longstanding childhood	--	Kappa Pearson's *r*	2 weeks	item level Kappa for subsample (N = 31) .79
Brief Daily Hassles Scale (Wright et al., 2010)	Frequency of events - Hassles from parents - Hassles from peers/others	.86 .83	--	--	--
Survey of Experiences (Allen et al., 2012)	Frequency/Impact of events Negative, Positive Events	--	Kappa Pearson's *r* Spearman's r	1 week 1 week	56% (Child) / 61% (Parent) Total frequency: .75 (Child) / .75 (Parent) Negative frequency: .28 (<12yrs) / .66 (≥ 12yrs) Positive frequency: .52 (< 12yrs) / .77 (≥ 12yrs) Negative impact: .59 (<12yrs) / .81 (≥ 12yrs) Positive impact: .35 (<12yrs) / .68 (≥ 12yrs)
Hispanic Stress Inventory - Adolescent (Cervantes et al., 2012)	Impact of events 8 subscales: - Family economic stress - Culture and Education stress - Acculturation stress - Immigration-related stress - Discrimination stress - Family immigration stress - Community/gang stress - Family/drug-related stress	.92 .85 .84 .82 .84 .78 .77 .69 .64	--	--	--

Structural Validity

Fourteen of the 21 included studies examined structural validity. Methodological quality ratings varied from 'poor' to 'excellent' (Column 5, Table 3). 'Poor' ratings were generally due to inadequate sample size (6 studies). Specifically, the COSMIN checklist recommends a minimum of seven times the number of items and over 100 participants for 'Excellent' quality with samples fewer than five times the number of items considered 'Poor' (Mokkink, Terwee et al. 2010). Samples ranged from as little as less than one times the number of items (Lewinsohn and Talkington 1979) to 24 times the number of items (Shahar, Henrich et al. 2003).

Of the 14 studies examining structural validity, 12 used either Principal Components Analyses (PCA) or Exploratory Factor Analysis (EFA) and 2 used Confirmatory Factor Analysis (CFA) – see Table 5 for summary of structural analyses. One study (Wright, Creed et al. 2010) initially identified the factor structure of the Brief Daily Hassles Scale using EFA and then replicated this using CFA in an independent but similar sample.

It is debated whether EFA or CFA is more appropriate for assessing structural validity. EFA (and PCA) are typically recommended when there are no *a priori* hypotheses about the underlying factors of a measure, whereas CFA compares specific hypothesised models of latent factors that underlie the data. In both cases a range of statistical 'fit' indices are used to identify the model that best fits the data. However, there is also debate regarding the most appropriate fit indices in each case (Bentler 2007). An appraisal of factor analytic methodology is beyond the scope of this review, but methodological details are given in Table 5 for descriptive purposes.

Using factor analysis, studies varied widely in the number and specific nature of subscales proposed. Some proposed broad scales of positive and negative events (e.g. Kanner, Feldman et al. 1987) whilst others described more specific life domains (e.g. Byrne, Davenport et al. 2007). Longer measures tended to identify more subscales.

Table 5. Structural Validity: Factorial analyses of life event measures

	Age Group (years)	Sample Size (n)	No. of items	Analytic Approach Used	Fit statistics	Factors identified (% variance explained)
Social Readjustment Rating Questionnaire (Coddington, 1972)	--	--	--	--	--	--
Unpleasant Events Schedule (Lewinsohn & Talkington, 1978)	15-??	76 (depressed) 61 (subclinical) 74 (controls)	320	PCA	49% variance explained for frequency, 51% for impact	1. social isolation 2. Marital discord 3. Inconveniences 4. Work hassle 5. Negative social interaction
Children's Own Perceptions and Experiences of Stressors (Colton, 1985)	8-11	181	60	EFA	44.5% variance explained	1. Feelings of isolation/rejection* 2. Major life events 3. Family disruptions 4. Cognitive Overload 5. Financial concerns 6. Step-families 7. School problems
Adolescent Hassles Inventory (Bobo et al., 1986)	6th Grade	246	68 50 (after PCA)	PCA	variance explained not reported	1. School (41) 2. Peers (12) 3. Future (8) 4. Drugs (7) 5. Parents (6) 6. Work (5) 7. Bother (4) 8. Money (4)
Adolescent Perceived Events Scale (Compas et al., 1987)	12-20			--		
Children's Hassles and Uplifts Scale (Kanner et al., 1987)	6th Grade	232	25 hassles 25 uplifts	PCA, Varimax (limited to 2 factors)	variance explained not reported	1. Hassles 2. Uplifts
Life Events and Coping Inventory (Dise-Lewis, 1988)	11-14	502	125 life events	PCA	40% variance explained, 76% of items	1 x life event scale

Adolescent Stress Questionnaire (Bagley, 1992)	14-16	369	68	PCA, Oblique	50.7% var explained	1. Relationship problems (18.4) 2. Abuse at home: drug/alcohol (8.1) 3. Scholastic/Career problems (6.5) 4. Heath/Personal concerns (females) (5.5) 5. Peer pressure (males) (4.2) 6. Material/Family problems (males) (4.1) 7. Loneliness/Isolation (females) (3.9)
Daily Life Stressors Scale (Kearney et al., 1993)	7-15	567	30	--		
Chinese Adolescent Life Event Scale (Cheng, 1997				--		
Urban Hassles Scale (Miller et al., 2002)	14-19	131	12 9 (after item analysis) *	PCA, Oblimin Rasch model (item level)	'infit', 'outfit' values from Rasch model given	1 x Urban Hassles Scale
Brief Adolescent Life Event Scale (Shahar et al., 2003)	12-15	895	36	CFA (hierarchical model)	RMSEA = 0.08	2 dimensions: Positive/Negative 4 Domains (pos/neg): - Achievement - Interpersonal 12 Facets (pos/neg): - family events - friendship events - peer-related/extracurricular - school events - work events - health/physical appearance
Stressful Life Event Schedule (Williamson et al., 2003)						
Urban Hassles Scale (Revised) (Miller et al., 2005)	10-20	254	32 21 (after PCA)	PCA, Orthogonal	51% variance explained	1. Environmental conditions (26.45) 2. Interpersonal interactions (10.96) 3. Safety concerns (8.14) 4. Anticipatory Victimisation (5.83)
Adolescent Stress Questionnaire (modified) (Byrne et al., 2007)	13-18	856*	58	PCA, Oblique	not reported	1. home life (9.7) 2. school performance (7.1)

						3. school attendance (4.3) 4. romantic relationships (9) 5. peer pressure (8.5) 6. teacher interaction (10) 7. future uncertainty (6.7) 8. school-leisure conflict (9.1) 9. financial pressure (8.7) 10. adult responsibility (1.7)
Hassles Measure for Indian Adolescents (Mehrotra & Kumari, 2009)	15-20	272	28	--		
Adolescent Stress Questionnaire - Norwegian (Moknes et al., 2010)[1]	13-18	723	58	PCA, Oblique	not reported	1.Stress of teacher/adult interaction (35.6) 2. Stress of peer pressure (4.8) 3. Stress of home life (4.1) 4. Stress of adult responsibility (3.8) 5. Stress of romantic relationships (3.1) 6. Stress of school attendance (2.7) 7. Stress of school-leisure conflict (2.5) 8. Stress of school performance (2.1) 9. Stress of financial pressure (2.0)
Linkoping Youth Life Experience Scale (Nilsson et al., 2010)	15-19	188	41	--		
Brief Daily Hassles Scale (Wright et al., 2010)	M = 15.35, sd = .79 (EFA) M = 16.1, sd = .75 (CFA)	212 (EFA) 236 (CFA)	69 49 (after item analysis) * 14 (after EFA)	EFA, Varimax CFA	RMSEA = 0.07	1. Parents (EFA: 40.0 2. Friends and others (EFA: 13.4)
Survey of Experiences (Allen et al., 2012)				--		
Hispanic Stress Inventory - Adolescent (Cervantes et al., 2012)	10-20	1651	160 119 (after	EFA, promax	81.6% variance explained	1. Family Economic stress 2. Culture/Educational stress

[1] Subsequent CFA conducted with refined version (Moknes et al, 2011) which identified the following 7 subscales: Teacher/adult interaction, Peer pressure, Home Life, Romantic Relationships, School attendance, School-leisure conflict, School performance.

	item analysis) * 72 (after EFA)	3. Acculturation-gap stress 4. Immigration-related stress 5. Discrimination stress 6. Family immigration stress 7. Community/gang-related stress 8. Family and drug-related stress

Hypothesis testing

Hypothesis testing refers to the degree to which scores on the target measure are related to other instruments based on the assumption that the measure validly measures the target construct (Mokkink, Terwee et al. 2010). Of the included studies, 17 examined relationships between the life event measure and some symptom measure. Methodological quality ratings ranged from 'poor' to 'good' (Column 6, Table 3). Those rated as 'poor' (n = 3, shown in red) generally lost points because they did not state their hypotheses *a priori*.

Drawing general conclusions is difficult given the variation in hypotheses tested (Table 6). Whilst most studies examined correlations between life event scale totals (e.g. frequency, impact) and symptom measures, others also examined correlations with the various proposed subscales[2]. Most studies examined correlations with well-established measures of depression (n = 4), anxiety (n = 1) or both (n = 9). Others included measures of self-esteem (Byrne, Davenport et al. 2007, Moksnes, Byrne et al. 2010) or general symptom measures (Allen, Rapee et al. 2012). Furthermore, the specific measure of symptoms (e.g. BDI vs CDI; state vs. trait anxiety) and parent- or child-reports varied across studies. However, there was a general pattern of moderate positive correlations between negative life event scales and both depression (range =.08 - .72) and anxiety (range = .10 - .60) and moderate negative correlations with self-esteem related measures (e.g. perceived social competence, r = -.27; Kanner, Feldman et al. 1987). Two studies (Lewinsohn and Talkington 1979, Williamson, Birmaher et al. 2003) examined mean differences in the number of events experienced between symptomatic young people ('psychiatrically ill' and depressed, respectively) and controls, finding greater endorsement of negative life events in symptomatic groups.

[2] For simplicity, only correlations for totals are reported in Table 6

Table 6. Hypothesis Testing: Associations between life event measures and symptom measures

	Sample size	Test	Measures	Correlation		Notes
Social Readjustment Rating Questionnaire (Coddington, 1972)	--					
Unpleasant Events Schedule (Lewinsohn & Talkington, 1978)	58*	Pearson's *r* (Occurrence)	BDI	.13, ns (frequency) .26, p <.05 (aversiveness)		*subsample *Also examined mean differences between depressed/not depressed (ANCOVA) - not reported
Children's Own Perceptions and Experiences of Stressors (Colton, 1985)	--					
Adolescent Hassles Inventory (Bobo et al., 1986)	246	Pearson's r (Impact)	Peer Relations Scale (Hudson, 1982)	.47 (.53 males/.36 females) *		p values not reported
Adolescent Perceived Events Scale (Compas et al., 1987)	--					
Children's Hassles and Uplifts Scale (Kanner et al., 1987)	232	Pearson's *r* (Frequency)		**Hassles***	**Uplifts**	
			STAI-trait	.60	-.03 (freq)	
			CDI	.38	-.38	
			Weinberger Adjustment Inventory - distress,	.43	-.26	
			Weinberger Adjustment Inventory - restraint	-.23	.17	
			Friendship Support Scale	-.22	.27	
			Perceived Social Competence Scale	-.27	.29	
			General Self-worth Scale	-.32	.26	* all sig (p <.05)
Life Events and Coping Inventory (Dise-Lewis, 1988)	198*	Pearson's *r* (Frequency)	STAI-state STAI-trait CDI	.28* .45 .43		*n varied depending on missing data * all sig (p <.05) *subscale correlations given in paper

Adolescent Stress Questionnaire (Bagley, 1992)	369	Pearson's *r* (impact)	BDI Cooper Smith Self-esteem Inventory for adolescents Suicidal Ideation	.22* .32 .20	* all sig (p <.05) *subscale correlations given in paper
Daily Life Stressors Scale (Kearney et al. 1993)	80*	Pearson's *r* (impact of events)	CDI Hopelessness Scale for Children STAI-state/trait Nowicki-Strickland Loss of control Scale Piers-Harris Self-concept scale	.72* .51 .52/.60 .67 -.74	* all sig (p < .05) *Foster-care group only
Chinese Adolescent Life Event Scale (Cheng, 1997)	483*	Pearson's *r* (impact of events)	BDI (Chinese version) STAI-trait (Chinese version)	.37 .20	* all sig (p < .05) *unique sample
Urban Hassles Scale (Miller et al., 2002)	--				
Brief Adolescent Life Event Scale (Shahar et al., 2003)	895	Pearson's *r* (frequency)	CDI	-.17 (negative events) * .37 (positive events)	*all sig (p < .05) *subscale correlations given in paper
Stressful Life Event Schedule (interview) (Williamson et al., 2003)	30 psych. ill 30 controls (no disorder)	Mean differences (T-test)		No of events: 8.1 (psych) vs 4.9 (cont.); $t_{58} = 3.35$, $p < .001$ Impact: 18.2 (psych.) vs. 9.9 (cont.); $t_{58} = 3.47$, $p < .001$	
Urban Hassles Scale (Revised) (Miller et al., 2005)	254	Pearson's *r* (frequency)	Hopkins Symptom Checklist - anxiety Hopkins Symptom Checklist - depression Adolescent Symptom Inventory	.48 .47 .48	*all sig (p < .01) * subscale correlation given in paper
Adolescent Stress Questionnaire (modified) (Byrne et al., 2007)	1039	Pearson's r (impact)	STAI-state State depression (study specific scale) Rosenberg Self-esteem Scale	correlations between these and each of the 10 subscales given in paper	
Hassles Measure for Indian Adolescents (Mehrotra & Kumari, 2009)	272	Pearson's *r* (frequency)	Penn State Worry Questionnaire (Indian version) Subjective Well-being (PGI- Indian)	.36* -.38	*all sig (p <.05)

Adolescent Stress Questionnaire – Norwegian (Moknes et al., 2010)	1183	Pearson's r (impact)	STAI- state State depression (as with Byrne et al., 2007) Rosenberg Self-esteem Scale	correlations between these and each of the 10 subscales given in paper	
Linkoping Youth Life Experience Scale (Nilsson et al., 2010)	188	Pearson's r (occurrence)	HADS - anxiety HADS - depression Dissociation Questionnaire (Swedish)	.11 - .26* .08 - .22 .19 - .38	*magnitude of correlations ranges across 4 subscales (.11 / .08, ns, all others sig, p < .05)
Brief Daily Hassles Scale (Wright et al., 2010)	236*	Pearson's r (frequency)	Reynolds Adolescent Depression Scale Multi-dimensional Anxiety Scale for Children Single-item Life satisfaction Likert scale	parent- / peer-related hassles .34 / .35* .19 / -.36 -.40 / .13	*independent sample * p < .05
Child and Adolescent Survey of Experiences (Allen et al., 2012)	80	Pearson's r (frequency)*	Strengths & Difficulties Questionnaire - parent Spence Children's Anxiety Scale - parent	Parent- / Child-report .38* / .06 .34* / .10	*p < .05 * correlations with child report SDQ/SCAS and with impact of events also reported in paper
Hispanic Stress Inventory – Adolescent (Cervantes et al., 2012)	1651	Pearson's r (impact of events)	CDI Achenbach's Youth Symptom Report	.41* .49	* p < .05

Cross-cultural validity

Only one study examined cross-cultural validity. All other measures were developed for use within a specific population so cross-cultural validity was not deemed relevant[3]. Monknes and colleagues (2010) adapted the Adolescent Stress Questionnaire (Byrne, Davenport et al. 2007) for use with Norwegian adolescents. They described a four-step translation process in which the original measure was first translated from English to Norwegian by three independent translators and then discussed until consensus was reached before being back-translated into English by two further independent translators. The translated measure was piloted prior to use with the study sample. This study was rated 'excellent' under COSMIN criteria, providing a good example of the process of adapting life event measures for use with other samples. However, results of this study also indicated differences in the factor structure between the original and adapted versions, highlighting possible cultural differences in the way that life events are experienced and conceptualised.

Criterion Validity

Criterion validity refers to the extent the target measure relates to a 'gold standard' measure (Terwee, Bot et al. 2007). No gold standard measure of life events in adolescence exists so criteria were relaxed to include all studies comparing the target measure with some other relevant life events measure, resulting in 4 relevant papers. COSMIN methodological quality ratings were all 'fair' (final column, Table 3).

Each study varied in its methods of assessing criterion validity (Table 7). For example, Compas (1987) compared self-reports on the Adolescent Perceived Events Scale with reports from the participant's college roommate in an older subset of their sample, whilst others compared child- and parent-report (Williamson et al., 2003; Allen et al., 2012), and one study compared the magnitude of stress-symptoms correlations between the Chinese Adolescent Life Event Scale and two other translated life event measures, but did not report the correlations between the life event measures (Cheng 1997).

The most comprehensive test of criterion validity was reported by Williamson and colleagues (2003). They not only compared self- and maternal-reports on their Stressful Life Event Schedule (SLES) interview but also compared responses on the SLES with existing checklist (Life Event Checklist (LEC); Johnson and McCutcheon 1980) and interview (Life

[3] Cheng (1997) also included Coddington's (1972) and Johnson & McCutcheon's (1980) measures translated into Chinese but did not examine the cross-cultural validity of these measures.

Events and Difficulties Schedule (LEDS); Monck and Dobbs 1985)[4] measures. This revealed 'near perfect' agreement (ICC = .81) between child- and parent-report on the SLES and 'substantial' agreement between parent-reports of the frequency of events on the SLES and LEC (ICC= .80) and child-reports of frequency on the SLES and LEDS interview (Kappa = .77).

Table 7. Studies examining criterion validity

	Criterion measure used	Findings
Adolescent Perceived Events Scale (Compas et al., 1987)	Self-report compared to responses of person in 'close relationship' i.e. college roommates	Subset of sample (18-20 years old) Found 82% agreement in occurrence of life events, 91% for frequency and 90% for impact
Stressful Life Event Schedule (interview) (Williamson et al., 2003)	Compared interview measure with alternative interview measure (Life Event and Difficulty Schedule; Brown et al., 1973)	Kappa = .77 (95% CI, .53-1.0)
	Compared child-report with parent-report on interview	ICC = .81 (95% CI, .68-.89)
Child and Adolescent Survey of Experiences (Allen et al., 2012)	Compared mother and child agreement on CASE checklist	60% agreement on the occurrence of life events
	Compared maternal-report on target measure (CASE checklist) and interview measure (PACE)	ICC=.62 for negative event frequency
Chinese Adolescent Life Event Scale (Cheng, 1997)	Compared correlations between target measure and anxiety measure with correlations between two other translated life event measures and depression and anxiety measures.	Similar magnitude stress-symptoms correlations (.23-.37 for depression, .16-.20 for anxiety) taken as evidence for criterion validity

Discussion

Summary

This systematic literature review aimed to provide an updated appraisal of measures of stressful life events in young people with a specific focus on the quality of the methodology used to develop such measures and the assessment of their psychometric properties.

A total of 21 unique measures of stressful life events met inclusion criteria. These largely consisted of self-report checklists of various life events developed in general population samples. The measures varied considerably in length, time frame captured, response scale

[4] This measure was not included in the current systematic review as full-text was unavailable

used, and the number and nature of subscales defined. The age ranges also varied across the included studies.

A range of psychometric properties were assessed. Most studies described measure development, allowing appraisal of content validity; most also examined the internal consistency of the developed measure and undertook a degree of hypothesis testing. Test-retest reliability, structural and criterion validity were less frequently examined and only one study examined cross-cultural validity. Studies varied considerably in the methodological quality of their assessment of each psychometric property. According to the COSMIN checklist, more recent studies tended to be of higher quality than older studies; perhaps reflecting increased emphasis on assessing psychometric properties of psychological measures.

Psychometric properties of life event measures

Content validity

Content validity is concerned with the relevance and comprehensiveness of a measure in capturing the target construct. The COSMIN standards identify 'bottom-up' methods, involving people from the target population in measure development, as more valid than 'top down' methods such as researcher-derived measures (Mokkink, Terwee et al. 2010). Of the studies describing development of the measure, only seven involved young people in either the development of the measure or in piloting its relevance and most studies only used young people to pilot measures derived 'top-down' by researchers.

Future research would benefit from more thorough consideration of the content validity of life event measures in young people. It is essential that the process of measure development is clearly described in all life events research, and new measures should be piloted with young people sampled from the target populations to enable other researchers to make judgements regarding the relevance of the scale to addressing their specific research questions. Whilst this creates a considerable time-consuming and research resource burden, it presents an important first step in developing a valid measure of young people's experiences and should be demonstrated in the development of all new life event measures.

Internal consistency

Over 40% of included studies did not report the internal consistency for the target life event measure. This is problematic because internal consistency is directly related to statistical

power to detect significant effects, so the effect sizes obtained in studies using scales with unknown internal consistency should be interpreted with caution.

Of studies assessing internal consistency, all scale totals exceeded the threshold for acceptable internal consistency. However, the internal consistency of constituent subscales (where reported) were typically lower with some in the unacceptably low range. Lower internal consistency estimates are common when scales contain few items (Streiner 2003). This highlights the importance of generating subscales that contain adequate relevant items to reliably capture that life domain. However, it is also possible to obtain lower internal consistency for life event measures because, unlike symptom measures such as those for depression, endorsing one item (life event) within a subscale does not necessarily mean an individual is likely to have experienced another life event within that scale, for example getting a 'bad school test result' and 'losing your homework' might both be included in the 'school events' subscale but the likelihood of an individual experiencing both events may not necessarily be more likely than chance.

It is important to note that estimates of internal consistency are not an inherent property of a measure and instead are reliant on specific sample characteristics. As recommended in guidelines produced by the APA task force on statistical inference (Wilkinson 1999), reliability coefficients should be reported even when psychometrics is not the focus of the research. As a result, it is essential that future studies examine internal consistency of their life event measure as a minimum standard, even if using an existing life event measure.

Test-retest reliability

Test-retest reliability refers to the stability of the measure across time. However, there is debate regarding the most appropriate method of assessing test-retest reliability of life event measures. Whilst most papers assessed agreement of total scores over time using Pearson's correlations between total scores, others examined item-level agreement. These result in different interpretations of the stability of life event measures with the former examining stability of the mean number (if frequency response scales are used) or impact (if severity scales are used) of events across time whilst the latter assesses the relative stability of specific items over time. In the absence of best-practice guidelines, future research may benefit from using both methods to comprehensively assess change in response over time.

The optimal interval between measurements for testing stability also needs to be considered. The interval between repeated administrations should be sufficient enough to prevent learning or carry-over effects but short enough to avoid problems with recall of

events from memory. This is further complicated in measures of life events which vary in the types of events captured. Measures of major events occurring across longer time periods may require adequate reliability across longer periods of time than daily hassles measures which are, by definition, expected to more changeable across time. As a result, subjective judgement is needed when selecting the retest interval for assessing test-retest reliability of life event measures.

Structural validity

Studies examining structural validity generally demonstrated poor methodological quality because of insufficient sample sizes. The COSMIN checklist proposes that sample sizes exceed seven times the number of items and more than 100 participants for 'Excellent' quality, placing considerable burden on resources to recruit large samples, especially for longer measures. However, if subscales from these measures are to be used to further our understanding of the complex interplay between different domains of life experiences and mental health outcomes, future studies should endeavour to adequately assess the underlying structure of the target measure first.

As discussed, there is considerable debate regarding the best practice for structural analyses, leading to diversity in methods used across studies. This makes it difficult to draw comparisons between studies regarding the underlying structure of different measures of life events in relation to different conceptualisations of stressful life events (Compas, 1987). However, general guidelines recommend that the factor structure of any new measure should first be examined using Exploratory Factor Analysis (EFA) which does not make any *a priori* hypotheses regarding its underlying structure and then researchers should seek to replicate this structure using Confirmatory Factor Analysis (CFA) in independent but similar samples. This approach was used in only one of the included studies. However, since only the original development papers were included in the current review, it is possible that subsequent factor analytic studies of some of the included measures exist.

Finally, as with statistical estimate of internal consistency, there are theoretical issues with using statistical analyses to determine structural validity of life event measures. Whilst items within subscales may be semantically linked (e.g. reflecting the same life domain) they do not necessarily co-occur and thus inter-item correlations could plausibly be near-zero so would not cluster together on factor analysis.

Hypothesis testing

The term hypothesis testing was used in current review to mean any study that examined associations between the target life event measure and some symptom measure; most commonly a self- or parent-reported measure of depression or anxiety but the specific measure used varied considerably across studies. As expected, negative event subscales generally showed moderate positive correlations with negative emotion symptom measures and significant negative associations with measures of well-being, with positive life event scales showing the inverse pattern. Although not described here, studies also varied regarding the extent to which they explored specificity in associations between life event subscales and symptom measures.

The discrepancy in symptom measures used across studies makes it difficult to draw conclusions about the relationship between life experiences and mental health outcomes. This further highlights the need to report the basic psychometric properties (internal consistency, test-retest reliability) in each specific study. Without this, it is difficult to conclude with confidence, particularly at the subscale level, whether mixed results across studies reflect meaningful difference between different aspects of the life event construct or are due to measurement error.

Criterion validity

The degree of convergence between the new life event measure and a 'gold standard' measure was rarely examined in the included studies. This may, in part be because no 'gold standard' exists. However, very few studies examined convergence with any other life event measure. The methodology for assessing criterion validity varied significantly with some papers directly comparing agreement between total scores from multiple measures whilst others compared the magnitude of stress-symptom correlations between measures. An interesting avenue for future research would be to include multiple measures of life events in a single sample of young people. This would allow comparison of the degree of association between them as well as whether they similarly predict hypothesised outcome measures, such as depression and anxiety.

Cross-cultural validity

Only one study included in the current review formally assessed its cross-cultural valdiity. As mentioned, this is generally because measures have been developed with specific cultural samples in mind. Life event measures may be culturally sensitive. Certain events may be experienced more/less frequently in different cultures and similalry, the perceived impact

of these events on the individual may also vary depending on cultural norms. The nature of cross-cultural variation in young people's experience of stressful life events is not well understood. Although outside the scope of the current review, this presents an interesting avenue for future research.

Strengths and Limitations

The main strength of the current review is the use of a standardised critical appraisal tool specifically designed to assess the methodological quality of studies examining psychometric properties (COSMIN; Mokkink, Prinsen et al. 2016) and the use of two independent raters. The high level of agreement between the raters suggests reliable assessment of study quality of each psychometric property assessed. However, the stringent criteria for assessing each psychometric property in the COSMIN checklist and the "worst score counts" approach, which weights all criteria equally when assessing overall quality, may result in lower estimates due to minor study flaws which may not have the same implications for interpretation of the quality of the measure.

Another major strength is the use of systematic and replicable search strategies across three established databases following PRISMA guidelines. This is an improvement on Blount and colleagues (2007) review, which relied on society members' endorsement of measures and identified only one measure of stress, Kanner's Children's Hassles and Uplifts Scale (1987), as 'well-established'. The current systematic review also included this measure, and a more thorough examination of its psychometric properties identified that whilst several properties were examined, the quality of these (using the COSMIN checklist) was generally in the poor to fair range.

The current systematic review was however limited. Whilst using the COSMIN checklist allowed for systematic evaluation of each measure's psychometric properties, the current review highlights some fundamental limitations of applying standard psychometric methods to life event measures. In particular, statistical estimates of internal consistency (Cronbach's alpha) and structural validity (factor analysis) may not be valid with life event measures where subscales are formed based on contextual similarities (e.g. relating to particular life domains) rather than on increased chances of co-occurrence that statistical psychometric methods rely on. More specifically, endorsing one life event item within a subscale does not necessarily mean an individual is likely to have experienced another life event within that subscale despite them being from similar life domains. This serves to reduce correlations between items and thus overall estimates of internal consistency and structural

validity. It may be more useful to focus on consensus-based approaches of assessing content validity. For example, this could involve asking samples of young people from the target population to generate the initial item pool, and piloting of new measures in additional samples of young people to ascertain the relevance and comprehensiveness of items included.

Other limitations were necessary to ensure a feasible number of studies for comprehensive review and synthesis. Studies were only included if they referred to 'mental health' symptoms and if only in young people aged less than 18 years. Whilst these restrictions were necessary to ensure a homogenous group of studies, measures developed in studies of other forms of psychopathology or in other research fields such as physical health may have been excluded. Only studies describing the original development of the measure used were included. It is generally good practice to assess and report the psychometric properties of a new measure when it is first used in a study however, it is possible that this restriction meant that subsequent studies examining the psychometric properties of the target measure (e.g. its structural or cross-cultural validity) were missed. As a result, it is recommended that researchers interested in using a measure detailed here conduct follow-up literature searches to identify further evaluations of its psychometric properties.

Conclusions and implications

A broad range of measures for assessing stressful life events in young people have been developed. These measures vary in focus and response format and studies describing their development vary in comprehensiveness and in the quality of psychometric properties assessment. The diversity of measures used in the included studies makes it difficult to make general recommendations regarding the 'best' or 'most useful' measure of stressful life events in young people. Instead several principles for selecting measures have been identified and are proposed. Researchers should ideally use one of the many existing measures rather than developing their own and are encouraged to think carefully about the type of life events likely to be relevant in their specific studies; for example, major life events versus daily hassles or events occurring within specific contexts such as urban environments. As a minimum standard, researchers should ensure that there is adequate evidence for content validity with the target population; for example, by piloting the measure in a subsample of the target population, to assess relevance, and by reporting basic psychometrics such as internal consistency. Using already established measures will reduce methodological differences across studies making it easier to draw conclusions regarding the interplay between stress and mental health outcomes in young people

The current review highlights several gaps in the literature regarding the measurement of stressful life events in young people. Most measures focused on self-report checklists used with children from middle childhood and beyond with no measures capturing the experiences of younger children. This presents one possible avenue for future research. In doing so it will be important to consider multiple informants, as previous research has indicated low convergence between parent- and child- (Rende and Plomin 1991, Johnston, Steele et al. 2003) and parent- and adolescent-reports (Kushner and Tackett 2017). Similarly, most measures were developed with samples drawn from the general population. Different clinical populations are likely to experience diverse life stressors and therefore existing measures may not be relevant or perform in the same way in samples of young people with different clinical characteristics. The application of existing measures with clinical samples with specific assessment of their psychometric properties (especially content and structural validity) presents another important avenue for future research.

This review also highlights the lack of consistency in how measures of stressful life events are developed with some examining psychometric properties more rigorously than others. It is important that all researchers take a critical approach to developing these measures by detailed examination of the full range of their psychometric properties. Fellow researchers can then make informed decisions when selecting a life event measure to address their research questions.

The COSMIN critical appraisal tool may be helpful for both designing a measure and reporting on the various stages of its development. However, the limitations of applying standard psychometric evaluations to measures of life events warrants further discussion of the most appropriate methodologies of validating such measures in future research. Given difficulties with statistical assessment of psychometric properties of life event measures it may be more important to focus on semantically coherent stressful life event measures using consensus-based approaches.

In summary, whilst a diverse range of measures of stressful life events in young people exist, future research would benefit from a more thorough examination of the psychometric properties of existing and new measures using standardised methods within a single sample. This will enhance our understanding of the dynamic influence of stressful experiences on childhood outcomes.

References

Allen, J. L., Rapee, R. M., & Sandberg, S. (2012). Assessment of maternally reported life events in children and adolescents: a comparison of interview and checklist methods. *Journal of psychopathology and behavioral assessment, 34*(2), 204-215.

Bagley, C. (1992). Development of an Adolescent Stress Scale for Use by School Counsellors Construct Validity in Terms of Depression, Self-Esteem and Suicidal Ideation. *School Psychology International, 13*(1), 31-49.

Bailey, D., & Garralda, M. E. (1990). Life events: Children's reports. *Social psychiatry and psychiatric epidemiology, 25*(6), 283-288.

Bentler, P. M. (2007). On tests and indices for evaluating structural models. *Personality and Individual Differences, 42*(5), 825-829.

Blount, R. L., Simons, L. E., Devine, K. A., Jaaniste, T., Cohen, L. L., Chambers, C. T., & Hayutin, L. G. (2007). Evidence-based assessment of coping and stress in pediatric psychology. *Journal of pediatric psychology, 33*(9), 1021-1045.

Bobo, J. K., Gilchrist, L. D., Elmer, J. F., Snow, W. H., & Schinke, S. P. (1986). Hassles, role strain, and peer relations in young adolescents. *The Journal of Early Adolescence, 6*(4), 339-352.

Byrne, D. G., Davenport, S., & Mazanov, J. (2007). Profiles of adolescent stress: The development of the adolescent stress questionnaire (ASQ). *Journal of adolescence, 30*(3), 393-416.

Chance, M. (1982). CopingBehaviors That Mediate Between Life Events and Depression. *Arch Gen Psychiatry, 39*, 1386-1391.

Cheng, C. (1997). Assessment of major life events for Hong Kong adolescents: The Chinese adolescent life event scale. *American Journal of Community Psychology, 25*(1), 17-33.

Coddington, R. D. (1972). The significance of life events as etiologic factors in the diseases of children: I—A survey of professional workers. *Journal of Psychosomatic Research, 16*(1), 7-18.

Cohen, L. H., McGowan, J., Fooskas, S., & Rose, S. (1984). Positive life events and social support and the relationship between life stress and psychological disorder. *American Journal of Community Psychology, 12*(5), 567-587.

Cohen, S., Kessler, R. C., & Gordon, L. U. (1997). *Measuring stress: A guide for health and social scientists*: Oxford University Press on Demand.

Colton, J. A. (1985). Childhood stress: Perceptions of children and professionals. *Journal of psychopathology and behavioral assessment, 7*(2), 155-173.

Compas, B. E. (1987). Stress and life events during childhood and adolescence. *Clinical Psychology Review, 7*(3), 275-302.

Compas, B. E., Connor-Smith, J., & Jaser, S. S. (2004). Temperament, stress reactivity, and coping: Implications for depression in childhood and adolescence. *Journal of Clinical Child and Adolescent Psychology, 33*(1), 21-31.

Compas, B. E., Davis, G. E., Forsythe, C. J., & Wagner, B. M. (1987). Assessment of major and daily stressful events during adolescence: the Adolescent Perceived Events Scale. *Journal of consulting and clinical psychology, 55*(4), 534.

Copeland, W. E., Keeler, G., Angold, A., & Costello, E. J. (2007). Traumatic events and posttraumatic stress in childhood. *Archives of General Psychiatry, 64*(5), 577-584.

Dixon, W. A., & Reid, J. K. (2000). Positive Life Events as a Moderator of Stress-Related Depressive Symptoms. *Journal of Counseling & Development, 78*(3), 343-347.

Dohrenwend, B. P. (2006). Inventorying stressful life events as risk factors for psychopathology: Toward resolution of the problem of intracategory variability. *Psychological bulletin, 132*(3), 477.

Duggal, S., Malkoff-Schwartz, S., Birmaher, B., Anderson, B. P., Matty, M. K., Houck, P. R., . . . Frank, E. (2000). Assessment of life stress in adolescents: Self-report versus interview methods. *Journal of the American Academy of Child & Adolescent Psychiatry, 39*(4), 445-452.

Eley, T. C., & Stevenson, J. (2000). Specific life events and chronic experiences differentially associated with depression and anxiety in young twins. *Journal of abnormal child psychology, 28*(4), 383-394.

Finlay-Jones, R., & Brown, G. W. (1981). Types of stressful life event and the onset of anxiety and depressive disorders. *Psychological Medicine, 11*(4), 803-815.

Ge, X., Conger, R. D., & Elder Jr, G. H. (2001). Pubertal transition, stressful life events, and the emergence of gender differences in adolescent depressive symptoms. *Developmental psychology, 37*(3), 404.

Ge, X., Lorenz, F. O., Conger, R. D., Elder, G. H., & Simons, R. L. (1994). Trajectories of stressful life events and depressive symptoms during adolescence. *Developmental psychology, 30*(4), 467.

Grant, K. E., Compas, B. E., Stuhlmacher, A. F., Thurm, A. E., McMahon, S. D., & Halpert, J. A. (2003). Stressors and child and adolescent psychopathology: Moving from markers to mechanisms of risk. *Psychological bulletin, 129*(3), 447.

Grant, K. E., Compas, B. E., Thurm, A. E., McMahon, S. D., & Gipson, P. Y. (2004). Stressors and child and adolescent psychopathology: Measurement issues and prospective effects. *Journal of Clinical Child and Adolescent Psychology, 33*(2), 412-425.

Hamilton, J. L., Stange, J. P., Shapero, B. G., Connolly, S. L., Abramson, L. Y., & Alloy, L. B. (2013). Cognitive vulnerabilities as predictors of stress generation in early adolescence: pathway to depressive symptoms. *Journal of abnormal child psychology, 41*(7), 1027-1039.

Hammen, C., & Rudolph, K. D. (1996). Childhood depression. In J. Mash, Barkley, R. A. (Ed.), *Child Psychopathology*. New York: Guildford.

Hankin, B. L., Mermelstein, R., & Roesch, L. (2007). Sex differences in adolescent depression: Stress exposure and reactivity models. *Child development, 78*(1), 279-295.

Hovens, J. G., Wiersma, J. E., Giltay, E. J., Van Oppen, P., Spinhoven, P., Penninx, B. W., & Zitman, F. G. (2010). Childhood life events and childhood trauma in adult patients with depressive, anxiety and comorbid disorders vs. controls. *Acta Psychiatrica Scandinavica, 122*(1), 66-74.

Hu, L. t., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural equation modeling: a multidisciplinary journal, 6*(1), 1-55.

Johnson, J. H., & McCutcheon, S. M. (1980). Assessing life stress in older children and adolescents: preliminary findings with the life events checklist. In S. I. G. & S. C. D. (Eds.), *Stress and Anxiety*. Washington DC: Hemisphere.

Johnston, C. A., Steele, R. G., Herrera, E. A., & Phipps, S. (2003). Parent and child reporting of negative life events: discrepancy and agreement across pediatric samples. *J Pediatr Psychol, 28*(8), 579-588.

Kanner, A. D., Coyne, J. C., Schaefer, C., & Lazarus, R. S. (1981). Comparison of two modes of stress measurement: Daily hassles and uplifts versus major life events. *Journal of behavioral medicine, 4*(1), 1-39.

Kanner, A. D., Feldman, S. S., Weinberger, D. A., & Ford, M. E. (1987). Uplifts, hassles, and adaptational outcomes in early adolescents. *The Journal of Early Adolescence, 7*(4), 371-394.

Kearney, C. A., Drabman, R. S., & Beasley, J. F. (1993). The trials of childhood: The development, reliability, and validity of the Daily Life Stressors Scale. *Journal of Child and Family Studies, 2*(4), 371-388.

Kercher, A. J., Rapee, R. M., & Schniering, C. A. (2009). Neuroticism, life events and negative thoughts in the development of depression in adolescent girls. *Journal of abnormal child psychology, 37*(7), 903-915.

Kersten, P., Czuba, K., McPherson, K., Dudley, M., Elder, H., Tauroa, R., & Vandal, A. (2016). A systematic review of evidence for the psychometric properties of the Strengths and Difficulties Questionnaire. *International Journal of Behavioral Development, 40*(1), 64-75.

Kim, K. J., Conger, R. D., Elder Jr, G. H., & Lorenz, F. O. (2003). Reciprocal influences between stressful life events and adolescent internalizing and externalizing problems. *Child development, 74*(1), 127-143.

Kushner, S. C., & Tackett, J. L. (2017). Parent and adolescent agreement for reports of life stressors. *Assessment, 24*(2), 143-156.

Lazarus, R. S. (1990). Theory-based stress measurement. *Psychological inquiry, 1*(1), 3-13.

Lewinsohn, P. M., & Talkington, J. (1979). Studies on the measurement of unpleasant events and relations with depression. *Applied Psychological Measurement, 3*(1), 83-101.

Miller, D. B., & Townsend, A. (2005). Urban hassles as chronic stressors and adolescent mental health: The Urban Hassles Index. *Brief Treatment and Crisis Intervention, 5*(1), 85.

Miller, D. B., Webster, S. E., & MacIntosh, R. (2002). What's there and what's not: Measuring daily hassles in urban African American adolescents. *Research on Social Work Practice, 12*(3), 375-388.

Moher, D., Shamseer, L., Clarke, M., Ghersi, D., Liberati, A., Petticrew, M., . . . Stewart, L. A. (2015). Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015 statement. *Systematic Reviews, 4*(1), 1.

Mokkink, L. B., Prinsen, C. A., Bouter, L. M., de Vet, H. C., & Terwee, C. B. (2016). The COnsensus-based Standards for the selection of health Measurement INstruments (COSMIN) and how to select an outcome measurement instrument. *Brazilian journal of physical therapy*(AHEAD), 0-0.

Mokkink, L. B., Terwee, C. B., Patrick, D. L., Alonso, J., Stratford, P. W., Knol, D. L., . . . De Vet, H. C. (2010a). The COSMIN checklist for assessing the methodological quality of studies on measurement properties of health status measurement instruments: an international Delphi study. *Quality of life research, 19*(4), 539-549.

Mokkink, L. B., Terwee, C. B., Patrick, D. L., Alonso, J., Stratford, P. W., Knol, D. L., . . . de Vet, H. C. (2010b). The COSMIN study reached international consensus on taxonomy, terminology, and definitions of measurement properties for health-related patient-reported outcomes. *Journal of clinical epidemiology, 63*(7), 737-745.

Moksnes, U. K., Byrne, D. G., Mazanov, J., & Espnes, G. A. (2010). Adolescent stress: evaluation of the factor structure of the adolescent stress questionnaire (ASQ-N). *Scandinavian Journal of Psychology, 51*(3), 203-209.

Monck, E., & Dobbs, R. (1985). Measuring life events in an adolescent population: methodological issues and related findings. *Psychol Med, 15*(4), 841-850.

Nilsson, D., Gustafsson, P. E., Larsson, J. L., & Svedin, C. G. (2010). Evaluation of the Linköping youth life experience scale. *The Journal of nervous and mental disease, 198*(10), 768-774.

Platt, R., Williams, S. R., & Ginsburg, G. S. (2016). Stressful Life Events and Child Anxiety: Examining Parent and Child Mediators. *Child Psychiatry Hum Dev, 47*(1), 23-34. doi:10.1007/s10578-015-0540-4

Rende, R. D., & Plomin, R. (1991). Child and parent perceptions of the upsettingness of major life events. *Journal of Child Psychology and Psychiatry, 32*(4), 627-633.

Rudolph, K. D., & Hammen, C. (1999). Age and gender as determinants of stress exposure, generation, and reactions in youngsters: A transactional perspective. *Child development, 70*(3), 660-677.

Rutter, M. (1981). Stress, coping and development: Some issues and some questions. *Journal of Child Psychology and Psychiatry, 22*(4), 323-356.

Seiffge-Krenke, I. (2000). Causal links between stressful events, coping style, and adolescent symptomatology. *Journal of adolescence, 23*(6), 675-691.

Shahar, G., Henrich, C. C., Reiner, I. C., & Little, T. D. (2003). Development and initial validation of the brief adolescent life event scale (BALES). *Anxiety, Stress & Coping, 16*(1), 119-128.

Stikkelbroek, Y., Bodden, D. H. M., Kleinjan, M., Reijnders, M., & van Baar, A. L. (2016). Adolescent Depression and Negative Life Events, the Mediating Role of Cognitive Emotion Regulation. *PLoS One, 11*(8). doi:10.1371/journal.pone.0161062

Streiner, D. L. (2003). Starting at the beginning: an introduction to coefficient alpha and internal consistency. *Journal of personality assessment, 80*(1), 99-103.

Terwee, C. B., Bot, S. D., de Boer, M. R., van der Windt, D. A., Knol, D. L., Dekker, J., . . . de Vet, H. C. (2007). Quality criteria were proposed for measurement properties of health status questionnaires. *J Clin Epidemiol, 60*(1), 34-42. doi:10.1016/j.jclinepi.2006.03.012

Thomson Reuters. (2016). EndNote X7. New York: Thomson Reuters.

Tosevski, D. L., & Milovancevic, M. P. (2006). Stressful life events and physical health. *Current Opinion in Psychiatry, 19*(2), 184-189.

Trickey, D., Siddaway, A. P., Meiser-Stedman, R., Serpell, L., & Field, A. P. (2012). A meta-analysis of risk factors for post-traumatic stress disorder in children and adolescents. *Clinical Psychology Review, 32*(2), 122-138.

Wilkinson, L. (1999). Statistical methods in psychology journals: Guidelines and explanations. *American psychologist, 54*(8), 594.

Williamson, D. E., Birmaher, B., Anderson, B. P., Al-Shabbout, M., & Ryan, N. D. (1995). Stressful life events in depressed adolescents: the role of dependent events during the depressive episode. *Journal of the American Academy of Child & Adolescent Psychiatry, 34*(5), 591-598.

Williamson, D. E., Birmaher, B., Ryan, N. D., Shiffrin, T. P., Lusky, J. A., Protopapa, J., . . . Brent, D. A. (2003). The stressful life events schedule for children and adolescents: development and validation. *Psychiatry research, 119*(3), 225-241.

Wright, M., Creed, P., & Zimmer-Gembeck, M. J. (2010). The development and initial validation of a brief daily hassles scale suitable for use with adolescents. *European Journal of Psychological Assessment*.

Appendices

Appendix 1. Psych Info Search Terms

1	exp "EXPERIENCES (EVENTS)"/ or exp LIFE EXPERIENCES/
2	(negative adj3 life).mp. [mp=title, abstract, heading word, table of contents, key concepts, original title, tests & measures]
3	(advers* adj3 event*).mp. [mp=title, abstract, heading word, table of contents, key concepts, original title, tests & measures]
4	(advers* adj3 experience*).mp. [mp=title, abstract, heading word, table of contents, key concepts, original title, tests & measures]
5	stress*.mp.
6	hassle*.mp.
7	life experience*.mp.
8	exp Stress/
9	(stress* adj3 experience*).mp. [mp=title, abstract, heading word, table of contents, key concepts, original title, tests & measures]
10	(stress* adj3 event*).mp. [mp=title, abstract, heading word, table of contents, key concepts, original title, tests & measures]
11	(stress* adj3 life).mp. [mp=title, abstract, heading word, table of contents, key concepts, original title, tests & measures]
12	(negative adj3 event*).mp. [mp=title, abstract, heading word, table of contents, key concepts, original title, tests & measures]
13	(negative adj3 experience*).mp. [mp=title, abstract, heading word, table of contents, key concepts, original title, tests & measures]
14	1 or 2 or 3 or 4 or 5 or 6 or 7 or 8 or 9 or 10 or 11 or 12 or 13
15	exp psychometrics/
16	exp test reliability/
17	exp test validity/
18	exp statistical reliability/
19	exp statistical validity/
20	internal consistency.mp
21	reliab*.mp. [mp=title, abstract, heading word, table of contents, key concepts, original title, tests & measures]
22	valid*.mp. [mp=title, abstract, heading word, table of contents, key concepts, original title, tests & measures]
23	psychometric*.mp. [mp=title, abstract, heading word, table of contents, key concepts, original title, tests & measures]
24	exp test standardization/
25	standardi$*.mp. [mp=title, abstract, heading word, table of contents, key concepts, original title, tests & measures]
26	15 or 16 or 17 or 18 or 19 or 20 or 21 or 22 or 23 or 24 or 25
27	exp Major Depression/ or exp "Depression (Emotion)"/ or exp Beck Depression Inventory/
29	depress*.mp.
29	low mood.mp.
30	(psych* adj3 symptom*).mp. [mp=title, abstract, heading word, table of contents, key concepts, original title, tests & measures]
31	(mental health adj3 (symptom* or outcome*)).mp. [mp=title, abstract, heading word, table of contents, key concepts, original title, tests & measures]
32	exp Well Being/ or well$being.mp.
33	27 or 28 or 29 or 30 or 31 or 32
34	14 and 26 and 33
35	limit 34 to (100 childhood <birth to age 12 yrs> or 200 adolescence <age 13 to 17 yrs>)

Comparing measures of stressful life events in young people

Abstract

Introduction: Stressful life events play a central role in many theoretical models of developmental psychopathology. However, such events have been conceptualised and measured in different ways across studies, complicating interpretation of research findings. Reviews of stressful life event measures used with young people highlight concerns regarding the methodological rigour with which these measures have been developed and the psychometric properties of these measures has rarely been examined.

Methods: This study systematically examined the psychometric properties of three existing measures of stressful life events in a large sample of adolescents to aid researchers in the selection of appropriate measures and to highlight effective methods for assessing their psychometric properties to guide future research. Internal consistency, test-retest reliability, structural and criterion validity were examined using the Consensus-based Standards for the Selection of health Measurement Instruments (COSMIN).

Results: Whilst all three measures demonstrated satisfactory internal consistency, test-retest reliability was inadequate. Furthermore, factor analyses revealed questionable structural validity for two of the three measures, and there was a lack of specificity in hypothesised associations between life event subscales and symptom measures of anxiety and depression.

Discussion: Analyses revealed issues with the psychometric properties of all three measures of life events. Future research will benefit from greater methodological rigour in the development, and subsequent assessment of the psychometric properties of stressful life event measures in young people. The development of a clearer working definition of stressful life events would facilitate greater consistency in study methodologies which would enhance our understanding of the complex role of stressful life events in developmental psychopathology.

Introduction

Stressful life events and developmental psychopathology

Theoretical models of developmental psychopathology highlight the importance of stressful life events in the development and maintenance of psychological problems in childhood and adolescence (Rudolph, Hammen et al. 2000, Dodge and Pettit 2003, Pagano, Skodol et al. 2004). Childhood and adolescence are particularly important developmental periods to explore because of the potential negative impact of stressful experiences on subsequent psychological development and well-being (Rutter 1981). As a result, understanding the impact of stressful life events on the development of psychological disorders in children and adolescents would greatly assist the development of interventions that reduce exposure of young people to stressors or that enhance young people's abilities to manage stressful experiences.

A review of the extensive stress-symptoms literature in young people is beyond the scope of the current study. However, several key themes have been identified.

There is robust evidence for prospective associations between stressful life experiences and subsequent onset of psychological symptoms in young people, both for severe traumatic experiences (e.g. major accident/injury, childhood abuse; Copeland, Keeler et al. 2007, Hovens, Wiersma et al. 2010, Trickey, Siddaway et al. 2012), as well as the cumulative effects of multiple stressful life experiences (Grant, Compas et al. 2004). Although limited, research suggests that more minor daily hassles, such as arguing with peers, may be more associated with the subsequent development of depression than major life events such as parental illness (Kanner, Coyne et al. 1981, Compas, Davis et al. 1987, Seiffge-Krenke 2000); however, few studies in young people have compared both major life events and daily hassles within a single study (Compas, Davis et al. 1987).

Research demonstrating varying forms of specificity in associations between certain types of life events and internalising symptoms has also been published. For example, there is evidence that dependent events, those that are in some way related to characteristics of the individual, and particularly interpersonal dependent events (e.g. peer conflict), are more related to depression symptoms than independent events (e.g. parental divorce) in young people (Williamson, Birmaher et al. 1995, Kercher, Rapee et al. 2009, Hamilton, Stange et al. 2013). In adults, studies have also identified a possible protective role of positive life events (e.g. satisfying social interactions, pleasant activities) as a buffer against psychological problems (Cohen, McGowan et al. 1984, Dixon and Reid 2000)

Finally, research has identified numerous possible moderating and mediating factors that affect the relationship between stressful life events and psychopathology. For example, the stress reactivity hypothesis of depression that proposed that personal characteristics (e.g. gender, cognitive style) moderate reactivity to stress which subsequently predicted depression severity in youth (Compas, Connor-Smith et al. 2004, Hankin, Mermelstein et al. 2007). Research has also suggested possible reciprocal relationships between stressful life events and psychopathology, in that vulnerabilities for depression (prior depressive episode, depressive cognitive style) may actually cause subsequent experiences of stressful life events rather than resulting as a consequence of life stress (Hamilton, Stange et al. 2013).

The studies identified suggest complex and dynamic associations between stressful life events and psychological problems across childhood and adolescence. However, differences in the way stressful life events have been conceptualised and measured across studies make it difficult to compare research findings and draw confident conclusions regarding the role of stressful life events in the development of psychopathology in youth with confidence.

Conceptualisations and measurement of stressful life events

A theoretical review of conceptualisations and measurement of stressful life events in young people has been provided in the systematic review contained within the current doctoral thesis. However, a summary of the key points relevant to the current research project are provided here.

Stressful life events are broadly defined as environmental circumstances or conditions that cause individuals psychological or biological distress. Stressful life events have been conceptualised in several different ways. Some researchers favour "*subjective*" or transactional definitions of stress where stress is seen as the result of individuals' cognitive appraisals of the type and severity of distress caused by specific environmental events or conditions (Lazarus 1990). Others favour an "*objective*" definition of stressful life events as "*environmental events or chronic conditions that objectively threaten the physical or psychological health or well-being of individuals of a particular age and in a particular society*" (Grant, Compas et al. 2003).

Additionally, stressful life events have been conceptualised as 'chronic' or 'acute' (Compas 1987), see Figure 1. Chronic events are stressful life events that are either recurring or result in enduring environmental disadvantage. Acute events involve qualitative changes in an individual's current environmental conditions. Acute life events can be further

differentiated into specific events which exert independent effects (e.g. life transitions), or the cumulative effects of numerous stressful life events over a specific period, as with major life events and daily hassles. Major life events refer to significant events impacting an individual's functioning (e.g. death of a loved one), whilst daily hassles refer to distressing everyday events (e.g. peer conflict).

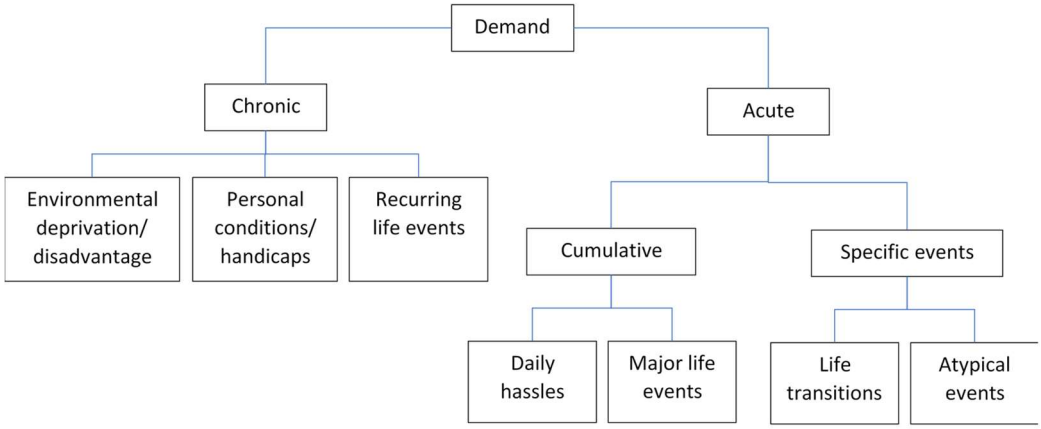


Figure 1. Conceptualisation of stressful life events (taken from Compas, 1987)

As well as differences in how stressful life events are conceptualised, researchers have used many different methods to measure stressful life experiences in youth. Some researchers have developed interview-based measures designed to provide objective assessment of the contextual threat associated with stressful events (e.g. Rudolph and Hammen 1999, Duggal, Malkoff-Schwartz et al. 2000), whilst others have used either self-, parent- or teacher-report checklists covering a range of typically occurring stressful life events (Brown, Smith et al. in prep). Few studies have looked at how different informants affect measurement of stressful life events with young people (Bailey and Garralda 1990, Rende and Plomin 1991, Johnston, Steele et al. 2003). However, those that have show that children (typically in middle childhood) generally report more negative events than their parents but less distress in relation to these events. Whilst parents may provide better information on stressors relating to the family context (e.g. parental health, wealth and relationships), children are likely to report on stressful experiences in the context of peer relationships more accurately (Rende and Plomin 1991). This is particularly important for measuring stressful life events in adolescents, who spend relatively more time away from their parents than younger children.

The stressful life event measures in studies in young people also vary in the response scales used; some ask respondents to indicate whether they have experienced a specific event

within a given time frame (occurrence) whilst others ask how often (frequency) and how severe (impact) the event was. Impact ratings fit with *subjective* conceptualisations of stressful life events whilst occurrence and frequency ratings are more in line with *objective* conceptualisations (Grant, Compas et al. 2003) and remove the possible confound with mediating variables such a cognitive appraisal style.

Finally, there is also variation in the types of stressful life events the measures capture. Some measures capture major life events (e.g. parental separation, illness or death) typically occurring over the previous 6-12 months (e.g. Coddington 1972, Williamson, Birmaher et al. 2003) whilst others measure daily hassles (e.g. arguments with friends or family) over brief intervals of weeks or months (e.g. Kanner, Feldman et al. 1987, Shahar, Henrich et al. 2003). Measures also vary in which life domains they capture, by using different subscales; for example, different settings (e.g. home, school) or contexts (e.g. peer, family, academic) or reflecting broader dependence-independence or interpersonal-non-interpersonal domains. It is also common for researchers to have designed bespoke measures of life events for use with specific samples (e.g. scales developed specifically for rural or urban populations (Miller and Townsend 2005)), resulting in great variability in the scope of the existing measures.

Reviews of life event measures in young people

In their review of the literature, Grant et al. (2004) identified over 500 studies using stressful life event measures with children and adolescents. Of these, 45% reported developing measures themselves with fewer than 10% using a well-validated measure and the remainder using one of approximately 50 measures available. The authors reported that the psychometric properties of these measures were rarely examined. Whilst this review highlighted the diversity of stressful life event measures used in studies in young people, the authors did not provide further details regarding the psychometric properties of specific measures as this was not the main aim of their review.

A taskforce of the American Psychological Association examined measures of both coping and stress in young people by asking members of their survey list to identify measures that they commonly used (Blount, Simons et al. 2007). Three measures of stressful life events were identified; the Children's Hassles and Uplifts Scale (CHUS; Kanner, Feldman et al. 1987), Coddington Life Event Scales (CLES; Coddington 1972) and Questionnaire on Resources and Stress (QRS; Holroyd 1974). Of these, only the CHUS was deemed "well-established" based on several criteria, including the measure being; presented in at least

two peer-review articles, described in sufficient details to allow replication, and assessed in terms of validity and reliability. Whilst the taskforce's review highlighted the lack of established measures of stressful life events in young people, the use of survey list members potentially introduced researcher bias in identifying preferred measures.

The analyses for these two reviews were conducted over 15 years ago and since their publication the study of the effects of stressful life events in young people has advanced.

To address the limitations of these studies, a comprehensive systematic literature review of stressful life events measures commonly used with young people, focussing specifically on the methodological rigour for their development and their psychometric properties, was conducted prior to commencing this research project (Brown, Smith et al. in prep). The methodological quality and psychometric properties of a variety of measures of stressful life events historically used in studies in young people were assessed using the Consensus-based Standards for the Selection of health Measurement Instruments (COSMIN; Mokkink, Prinsen et al. 2016). COSMIN is a taxonomy for assessing common psychometric properties (e.g. internal consistency, reliability, construct validity, criterion validity) in health-related patient outcome measures. A total of 21 unique measures of stressful life events with some degree of assessment of psychometric properties were identified. However, these varied considerably in their response format (e.g. checklist vs interview), focus (e.g. major life events vs. daily hassles) and specific domains captured (e.g. life areas vs. dependent-independent events). The measures also varied in the extent to which their psychometric properties had been considered during their development and subsequently assessed. Whilst the majority of studies described the development of the measurement, allowing assessment of content validity, and reported their internal consistency, very few studies examined their test-retest reliability, structural validity or their convergence with existing life event measures.

This systematic review highlighted the need for a more comprehensive, systematic evaluation of the psychometric properties and performance of stressful life event measures in youth. Full details of the review are given in the first part of this thesis.

Current study

The current study aimed to address one of the main recommendations from the systematic review; that is, to include multiple measures of stressful life events in a single study (Brown, Smith et al. in prep). This allowed direct comparison of the psychometric properties of similar measures without the confounds of different sample characteristics. Data for the

current study was collected as part of a broader research project exploring stress-reactivity and stress-generation hypotheses of depression in adolescents.

Measure selection

Three established life event measures were included; the Children's Hassles and Uplifts Scale (CHUS; Kanner, Feldman et al. 1987), the Brief Adolescent Life Event Scale (BALES; Shahar, Henrich et al. 2003) and Adolescent Life Event Questionnaire (ALEQ; Hankin and Abramson 2002) in a large sample of young people (aged 11-14 years) using the COSMIN standards (See Appendices 3-5 for full measures).

The CHUS was included because it has been frequently used over many years in life events research and was the most cited of the life event measures identified in our systematic review (115 citations; Web of Science, March 2018). Although Blount and colleagues (2007) identified the CHUS as 'well-established' in their literature review, using the COSMIN checklist, our recent systematic review identified quality ratings for the assessment of psychometric properties largely in the poor-to-fair range, including poor reporting of structural analyses and no examination of test-retest reliability (Brown, Smith et al. in prep).

The BALES was included as a more recent measure of daily hassles, with similar content to the CHUS but with fewer items. The developers aimed to address the practical need of many researchers who employ large multivariate assessment protocols by capturing multiple life event domains with minimal items (Shahar, Henrich et al. 2003). The BALES was also identified in our systematic review with similar poor-to-fair quality ratings to the CHUS, including poor content validity and no assessment of test-retest reliability.

Finally, the ALEQ was also selected. Despite the ALEQ being used in numerous recent studies, the psychometric properties have not been assessed in detail (although Hankin & Abramson, 2002 have reported internal consistency and test-retest reliability over 2 weeks for the measure). Inclusion of the ALEQ allows comparison of the psychometric properties of a measure of major life events (where items are endorsed over the previous 3-6 months) with the CHUS and BALES, both considered as daily hassles measures, covering the previous week. The ALEQ was also included to meet the studies broader aim of replicating previous studies examining the stress-generation hypothesis of depression in adolescence (Hamilton, Stange et al. 2013).

Study Aims

Aim 1: Including these three measures within a single study allows for a greater understanding of their psychometric properties and it is hoped that direct comparison of between them will facilitate appropriate selection of these measures in future research aiming to measure daily hassles or major life events in adolescence.

Aim 2: The analyses reported here also aim to highlight appropriate methods for assessing the psychometric properties of life event measures to inform future researchers when developing measures of life events.

Methods

Study design

The current study forms part of a larger multi-wave longitudinal project exploring the relationships between stressful experiences, cognitive vulnerabilities and depression in young people, in line with stress-reactivity and stress-generation hypotheses.

This study consisted of three stages of data collection using self-report questionnaires across a six-month interval (Appendix 1). Data collection began in the Spring term of the academic year and finished part-way through the summer term. At time one, baseline depression and anxiety symptoms, irritability, rumination and negative cognitive style were measured. Five months later, at time two, four, weekly measures of daily hassles using the two different life event measures were taken. Finally, at time three, retrospective reports of life events over the preceding 6 months were collected together with self-report questionnaires of depression symptoms and irritability.

At each time point, questionnaires were administered to whole classes during form time. Research assistants provided support to children as needed. At time one, data collection took approximately 45 minutes to complete. The brief measures at time two took around 15 minutes on each occasion and measures at time three took approximately 30 minutes to complete. Those requiring extra time completed measures with the support of research assistants at a time in the school day negotiated with the young person and their teacher.

It was hoped that the possible benefits of participating helped to mitigate the considerable time burden placed upon the young people. At an individual level, we hope that students found it interesting to complete questionnaires considering their mental health; something they may not have considered previously. Additionally, completion of depression and anxiety measures allowed identification of young people with elevated symptoms which

may have otherwise gone undetected and facilitated an opportunity for these young people to receive professional support with their mental health.

Ethical considerations

Ethical approval was granted by the Psychiatry, Nursing and Midwifery Research Ethics Subcommittee of King's College London (Ref no: HR-15/16-1919).

Given the sensitive nature of the questionnaires administered, it was possible that young people may have experienced some emotional distress when completing the measures. To manage this, all young people were informed at each stage of data collection that their participation was completely voluntary and that they could opt-out or withdraw their data at any point without penalty. Researchers remained present throughout all data collection to monitor for signs of distress and young people were invited to complete measures and/or speak individually with researchers at any stage of data collection should they wish.

Immediately following each stage of data collection, all questionnaires were scrutinised by the researchers. If any items indicating possible risk were endorsed by the young person or the young person scored above the recommended clinical cut-off for either anxiety or depression, they were approached discreetly by the research team prior to informing the pastoral lead in the school to ensure local safeguarding procedures were followed.

Various forms of follow-up support were made available to young people. Participants were encouraged to approach researchers at any point throughout data collection. Furthermore, the contact details of the researchers and their supervisors (both practicing clinical psychologist in fields of anxiety and depression) were given to staff members who were encouraged to contact to discuss any concerns arising from the research project. At total of 15 young people reported either 'very elevated' depression scores or endorsed the risk item on the CDI ("I want to kill myself") at either time one or time three. Unfortunately, data regarding whether these young people went on to receive any specific mental health support was not available from school staff.

Participants

Participants were recruited from a single secondary school in South London. All students in years 7 – 9 (aged 11 – 14 years) were invited to participate. This age range was selected to capture adolescents around the typical age of onset for first episode depression (Birmaher, Sermiento, Ryan, Perel, Sanabria et al., 1996) and in line with research indicating that stressful life events prospectively predict the onset of depression in young people (Grant et al., 2004).

Recruitment used an opt-out parental consent and young person verbal assent procedure. Researchers attended class assemblies and form times to outline the research project and answer any questions. Information sheets for the parents and young person were provided to students. Information sheets and opt-out consent forms were also posted to parents along with stamped return envelopes. Parents were given two weeks to return the opt-out form if they did not want their child to participate. Adolescents provided verbal assent at the beginning of data collection at time-points one and three.

The total sample consisted of 249 adolescents[5]; 130 males (52%), 102 females (41%) and 17 (7%) with missing gender information[6] aged between 12 years, 2 months and 14 years, 4 months at time-point one (M = 13 years, 2 months, SD = .59). Most participants identified themselves as Black (55.9%), followed by White (21.2%), Mixed (13.2%), Asian (5.7%) and other (4%). This was in line with the ethnic distribution of the wider school. At time-point three, the sample consisted of 209 adolescents (115 male, 91 female) with a mean age of 13 years, 8 months (SD = .61). The sample did not vary significantly in gender or ethnic distributions across time points (Appendix 2).

Stress Measures

Children's Hassles and Uplifts Scale (CHUS; Kanner, Feldman et al. 1987).

The CHUS consists of two subscales measuring the frequency and impact of daily hassles (irritating/distressing everyday events) and uplifts (pleasant/positive everyday events). The measure was derived from a pool of 74 items generated by young people using factor analysis to identify the two higher-order scales each consisting of 25 items; hassles (e.g. you lost something, someone at school teased you) and uplifts (e.g. you got a good mark at school, you had a nice time with your friend). The original CHUS asked respondents to indicate whether they had experienced each life event in the last month (*occurrence*) and then to rate how bad (hassles) or good (uplift) they were on a three-point scale (*intensity* ratings). Higher scores indicate more life events. See Appendix 3 for full scale.

[5] Data regarding the proportion of total possible sample recruited is not available. Sample sizes varied across time points due to who was present in class at the time of administering the measures. Additionally, participants could opt-out of the study at any time point and then had the option to opt back in at later time points meaning that participants might have missing data for some time points but not others. However, it is important to note that the sample did not differ significantly in gender or ethnic distributions across the different time points (Appendix 2).

[6] Demographic information was completed at times 1 and 3. As a result, gender information was missing from participants that opted-out of these time points but completed time 2.

The psychometric properties of the measure were originally assessed in a sample of 232 6th graders, aged 10-13 years. Internal consistency was good for both subscales (αs = .87 for both hassles and uplifts) and both were correlated in the expected directions with scores on the Children's Depression Inventory (CDI; *r*s = .38 hassles and -.38 uplifts). Hassles, but not uplifts, were also significantly positively associated with anxiety symptoms on the State Trait Anxiety Inventory-trait scale (*r* = .60). Structural validity of the CHUS was assessed using Principal Components Analysis which confirmed distinct hassles and uplifts scales. However, the variance explained by the model was not reported. Test-retest reliability and criterion validity was also not assessed.

Brief Adolescent Life events Scale (BALES; Shahar, Henrich et al. 2003)

The BALES consists of 36 items covering six life domains; family life, close friendships, peer relations, extracurricular activities, school performance, and health and appearance concerns. Items were derived by the authors from previous life event measures to represent both positive and negative events and to equally represent interpersonal and achievement-related/non-interpersonal events. Participants rate how *frequently* each event has occurred in the previous week; never (0), a little (1), some (2) or a lot (3). Items can be summed to calculate four subscales; Negative Interpersonal Events, Negative Achievement Events, Positive Interpersonal Events and Positive Achievement Events. Higher scores indicate more life events. See Appendix 4 for full scale.

The original development paper examined the psychometric properties of the measure in a large sample of 12-15-year-olds. In this study, the BALES demonstrated high internal consistency for both positive (α = .88) and negative (.87) scales with moderate internal consistency across subscales (.60 - .80). The positive and negative subscales were correlated as predicted with the Children's Depression inventory (CDI) (*r* = -.17 positive, *r* = .37 negative – correlations with subscales also reported). Structural validity was assessed using confirmatory factor analysis (CFA) which confirmed two higher-order dimensions reflecting positive and negative life events with 4 lower-order factors reflecting positive and negative interpersonal and non-interpersonal/achievement events. However, the proportion of variance explained by the model was not reported. Furthermore, test-retest reliability was not examined.

In the current study, participants were asked to indicate how frequently (never (0), a little (1), some (2), a lot (3)) they had experienced each event in the preceding week for both the CHUS and the BALES. Changing the CHUS response format to match the BALES allowed

direct comparison between them in our sample. This also removed the element of subjective interpretation regarding the salience of the events in the CHUS.

Adolescent Life events Questionnaire (ALEQ; Hankin & Abramson, 2002)

The ALEQ is a self-report checklist assessing 57 negative life events that commonly occur during adolescence, including familial (e.g. your parents divorced), peer (e.g. your friends didn't seem to understand you) and achievement (e.g. you did poorly on, or failed, a test) events. The ALEQ can be categorised into subscales representing interpersonal-dependent events (e.g. "you argued with a close friend"), interpersonal-independent events (e.g. "your close friend moved away") and non-interpersonal events (e.g. "you didn't complete the required homework"). See Appendix 5 for the full scale.

Hankin and Abramson (2003) reported high internal consistency (α = .94) and moderate test-retest reliability over two weeks (*r* = .65) in a sample of adolescents aged 13-18 years. However, the measure has not been subjected to factor analysis to ascertain its structural validity and the internal consistencies and reliabilities of the subscales have not been reported.

In the current study, adolescents completed the ALEQ at time-point three and indicated which life events had occurred since time-point one (six months previously) and how frequently they had occurred; never (0), rarely (1), sometimes (2), frequently (3) or always (4). Higher scores indicate more life events.

Symptom measures

Self-report measures of anxiety and depression symptoms were also included in the current study to examine correlations between specific life event measures and commonly-reported emotion symptoms in adolescence, in line with developmental models of psychopathology (Rudolph et al., 2000). The Children's Depression Inventory (CDI; Kovacs, 1982) and the Multidimensional Anxiety Scale for Children (MASC; March, Parker, Sullivan, Stallings & Conners, 1997) were selected because both represent well-established measures for use with adolescent samples, frequently used in both clinical practice and research. Furthermore, these measures were in line with the wider research project's aim to replicate previous research findings regarding stress reactivity and generation hypotheses of depression (Compas et al., 2004; Hankin et al., 2007).

Children's Depression Inventory (CDI; Kovacs 1982)

The CDI is 27-item self-report questionnaire assessing common symptoms of depression in youth between 8 and 14 years. For each item, respondents are given a group of three sentences and asked to choose the one that best describes their feelings in the past two weeks (e.g. "I am sad once in a while", "I am sad many times", "I am sad all the time"). Responses are coded 0 – 2 in line with severity and summed across items to give total depression scores. Higher scores indicate greater severity of depression symptoms. The CDI demonstrates adequate internal consistency (typically greater than .80) and test-retest reliability (typically around .80) in community samples (Kovacs, 1982; Saylor, Finch & Bennett, 1984). Internal consistencies in the current sample were .90 at time one and .88 at time two.

Multidimensional Anxiety Scale for Children (MASC; March et al. 1997)

The MASC is a 50-item self-report questionnaire measuring common anxiety symptoms in young people. Respondents rate how often they have experienced each item is true for them using a 4-point Likert scale; never (0), rarely (1), sometimes (2), often (3). Responses are summed across items to create total anxiety scores; higher scores indicating greater severity. The measure demonstrates high internal consistency (α = .90) and excellent test-retest reliability over 3 months (ICC = .93) in a community sample (March, Parker et al. 1997). The internal consistency in the current sample was .96.

Analyses

Data was coded and entered in SPSS (IBM Corp. 2016). Checks of the accuracy of data entry were completed by an independent researcher who selected a random subset of 10% of the questionnaires at each time point and manually checked data entry correctness. Scale and subscale totals were calculated for depression, anxiety and each life event measure and constituent subscales using individual mean replacement for each participant if less than 6% of items were missing. This meant that a maximum of 3 questions per questionnaire were pro-rated (e.g. 6% of the 50 -item CHUS = 3 items) with data excluded if participants were missing more than 6% of items, representing a conservative way of managing missing data. For example, 3.02% of time 1 total depression scores and 3.61% of time 2 total anxiety scores were missing due to exceeding this pro-rating threshold.

Differences in demographics and baseline variables between those that completed time one but did not complete either time two or three were compared using chi-square tests for frequency data (gender, ethnicity) and independent t-tests for continuous variables (age,

time-point one depression, time-point one anxiety). Completion at time two was defined as having data for at least three of the four weekly measures of life events and at time three it was defined as having completed at least one of the core variables at time three.

All analyses were conducted in SPSS (IBM Corp., 2016). Analyses examined each criterion outlined in the COSMIN checklist (Table 1) using best practice guidelines where possible.

Table 1. Definition of psychometric properties according to COSMIN (Mokkink et al., 2010)

	Definition	COSMIN quality criteria
Content validity	The degree to which the content of the instrument is an adequate reflection of the construct to be measured	1. All items refer to the relevant aspects of the construct to be measured 2. All items are relevant for the study population e.g. focus group (sample size >10) 3. All items relevant for the application of the measure 4. All items together comprehensively reflect the construct to be measured
Internal Consistency	The degree of interrelatedness between items	1. Scale consists of effect indicators i.e. based on a reflective model 2. Adequate sample size (≥ 100) 3. Unidimensionality checked using factor analysis 4. Internal consistency statistic calculated for each (sub)scale 5. Appropriate statistical methods; Cronbach's Aplha for continuous scores
Test-retest reliability	The extent to which scores on the same version of questionnaires for people who have not changed are the same for the repeated measurement over time	1. Adequate sample size (≥ 100) 2. At least to measurements available 3. Independent administrations 4. Participants stable between administrations and similar test conditions 5. Appropriate time interval 6. Appropriate statistical methods; intraclass correlations coefficient for continuous scores
Structural validity	The degree to which scores of an instrument are an adequate reflection of the dimensionality of the construct to be measured	1. Scale consists of effect indicators i.e. based on a reflective model 2. Adequate sample size (≥ 100 or 5-7 x no. of items but <100) 3. Exploratory or confirmatory factor analysis performed appropriately in view of existing information
Criterion/Convergent Validity	The correlation of the instrument with a "gold standard" criterion administered at the same time	1. Adequate sample size (≥100) 2. "Gold standard" criterion used 3. Appropriate statistical methods: Correlations used for continuous scores
Hypothesis testing	The degree to which scores of the instrument relate to	1. Adequate sample size (≥100)

	scores on other measures to which it should be related	2. Multiple hypotheses formulated a priori including expected direction and magnitude of effect 3. Adequate measurement properties of the comparator instrument(s) in a population similar to the study population 4. Appropriate statistical methods for the hypotheses to be tested

*COSMIN: **CO**nsensus-based **S**tandards for the selection of health **M**easurement **IN**struments*
Cross-cultural validity was not examined in the current study

Internal consistency*:* Internal consistency was assessed for each measure at the first administration using Cronbach's alpha (α) for totals and subscales. Values between 0.70 and 0.95 for subscales are considered acceptable (Terwee, Bot et al. 2007). Change in the mean number of events reported across the four x weekly measurements for each subscale on the daily hassles measures was examined using ANOVAs.

Test-retest reliability: Test-retest reliability was examined for both the CHUS and BALES using Intraclass correlations (ICCs) between responses on the first (time 2a) and fourth (time 2d) administration. ICCs were calculated rather than Pearson's correlation coefficient because they account for both interindividual (variation in the population) and intraindividual (measurement error) variation (Terwee, Bot et al. 2007). The reliability coefficient cut-off of .70 was used as recommended for samples greater than 50. The ALEQ was only administered once because of the six-month interval needed so it was not possible to examine its test-retest reliability in the current study.

Structural validity: The COSMIN quality criteria outline the "*appropriate use of exploratory or confirmatory factor analysis in light of existing information*" is used to examine structural validity (Table 1). However, there is much debate in the literature regarding what is appropriate and subsequently different methods are commonly used (Harrington 2008, Field 2013). The structural validity of the CHUS was originally assessed using principal components analysis (PCA) with a varimax (orthogonal) rotation and fixed to two factors (Kanner, Feldman et al. 1987). This identified two semantically relevant subscales reflecting hassles and uplifts, respectively. Conversely, the structural validity of the BALES was initially assessed using hierarchical confirmatory factor analysis (CFA) of subscales proposed by the authors to reflect 2-higher order dimensions of positive and negative events; each with two lower-order domains reflecting interpersonal and achievement-related events (Shahar, Henrich et al. 2003). However, the authors did not

compare this model to competing models (e.g. unidimensional or orthogonal models. Finally, the structural validity of the proposed subscales of the ALEQ has not been explored using factor analysis.

Since we had no consistent *a priori* theoretical model of the underlying structure of stressful life events, PCA was considered most appropriate in the current study. PCA is concerned with deconstructing the original data into factors whilst other forms of factor analysis impose theoretical or statistical assumptions on estimated data (Field 2013). PCA was conducted on the first administration of each life event measure. Using PCA for all measures included in the current study allows comparison of factor structures between them.

Kaiser-Meyer-Olkin's (KMO) measure of sampling adequacy (recommended cut off = .50) and Bartlett's test of sphericity were used to examine the basic sampling assumptions needed for PCA. Factors were extracted using the varimax rotation which assumes uncorrelated factors. The likely number of factors for each measure was identified using scree plots and eigenvalues greater than one (Field 2013). The percentage of variance explained by each factor is reported. Individual factors for each measure were then explored to assess their semantic relevance and compared to those proposed by the original authors.

Criterion validity: In the absence of a 'gold standard' measure of stressful life events in young people, criterion validity was assessed using Pearson's correlations to estimate the degree of convergence between the subscales of the three measures. Williams' tests were used as recommended by Weaver and Wuensch (2013) to compare the relative magnitude of correlations between subscales. It was hypothesised that:

1. the two daily hassles measures (BALES, CHUS) would be more strongly positively correlated with each other than either of these measures with the measure of major life events (ALEQ).
2. the negative interpersonal subscales on the BALES and ALEQ will be more strongly positively correlated than with non-interpersonal scales on these measures
3. the BALES negative achievement scale would be more strongly positively correlated with the ALEQ non-interpersonal scale than the other scales.

Hypothesis testing: Finally, Pearson's correlations were used to examine associations between life event measures and symptoms of depression and anxiety at time one and depression symptoms at time three[7]. It was hypothesised that:

1. the negative event subscales on all measures would be positively correlated with both anxiety and depression whilst positive event subscales would be negatively correlated with depression, in line with the stress-buffering hypothesis (Cohen, McGowan et al. 1984).
2. More specifically, subscales on each measure pertaining to negative dependent and interpersonal events would be more strongly correlated with depression than subscales pertaining to independent and non-interpersonal/achievement events, in line with previous research (Hamilton, Stange et al. 2013).
3. Finally, daily hassles (as measured on the CHUS and BALES) would be more strongly correlated with depression than major life events from the ALEQ, in line with research suggesting that the accumulation of everyday negative events is more predictive of concurrent and subsequent depression than major life events (Seiffge-Krenke 2000).

Results

Descriptive statistics

Descriptive statistics for all variables are shown in Table 2. Means from meta-analytic studies for anxiety and depression and from the original development papers for each life event measure are shown in the last column for comparison.

[7] Anxiety was not measured at time 3.

Table 2. Means (standard deviations) of study variables

CDI	Range 0 – 56		Time 1 9.53 (8.49)			Time 3 7.92 (7.11)	*Twenge et al (2002)* *8.98 - 10.06**
MASC	Range 0 – 150		Time 1 48.19 (26.99)			--	*Muris et al (2002)* *38.0 (18.8)*
BALES	Range	Time 2 a	Time 2 b	Time 2 c	Time 2 d	Mean across time	*Shahar et al (2003)*
Negative Interpersonal Experiences	0 – 27	3.92 (3.70)	3.02 (3.22)	2.75 (3.11)	2.67 (3.39)	3.04 (2.72)	16.8 (4.5)
Negative Achievement Events	0 – 27	4.77 (3.69)	3.89 (3.30)	3.50 (3.52)	3.35 (3.73)	3.82 (3.00)	18.6 (4.6)
Positive Interpersonal Events	0 – 27	10.94 (4.91)	9.37 (5.71)	8.96 (5.73)	8.03 (5.56)	9.47 (4.84)	7.3 (3.9)
Positive Achievement Events	0 – 27	14.13 (5.84)	12.42 (5.86)	11.20 (6.52)	10.23 (6.38	12.24 (5.33)	8.2 (4.6)
CHUS							
	Range	Time 2 a	Time 2 b	Time 2 c	Time 2 d	Mean across time	Kanner et al., (1987)
Hassles	0 – 75	15.51 (9.08)	13.34 (9.21)	11.52 (8.05)	10.97 (8.84)	12.95 (7.96)	12.00 (7.0)
Uplifts	0 – 75	38.25 (13.39)	35.17 (14.58)	32.92 (14.73)	31.45 (15.86)	34.08 (12.96)	17.5 (7.8)
ALEQ	Range					Time 3	Hankin & Abramson (2002)
Interpersonal-dependent	0 – 104					5.62 (5.21)	12.72 (5.46)
Interpersonal-independent	0 – 52					1.69 (1.94)	4.67 (2.14)
Non-interpersonal	0 – 40					3.36 (2.64)	4.70 (2.13)

**Standard deviation not reported. CDI = Children's Depression Inventory, MASC = Multi-dimensional Anxiety Scale for Children, BALES = Brief Adolescent Life Events Scale, CHUS = Children's Hassles and Uplifts Scale, ALEQ = Adolescent Life Event Questionnaire*

Note: sample sizes vary across measures and time points due to missing data

Descriptives from the measures original papers are shown in the final column for comparison

Depression and anxiety Mean depression scores were largely in line with meta-analytic findings for the CDI in children of the same age range; see final column of Table 2. Mean depression scores were significantly lower at time three than time one, *t* (150) = 2.35, *p* =.02; M_{time1} = 9.53, SD = 8.49, M_{time2} = 7.92, SD = 7.11. Females reported significantly more depression symptoms than males at time one, *t* (162) = -2.56, *p* = .01, $M_{females}$=11.35, SD = 9.48, M_{males}= 8.00, SD = 7.29, but not time three, *t* (193) = -1.59, *p* = .11, $M_{females}$ = 10.56, SD = 8.76, M_{males} = 7.67, SD = 6.98.

The mean anxiety score for the MASC at time one was 48.11, SD = 26.84. This is higher than that reported in a sample of 521, 12 to 18-year-olds (M = 38.00, SD = 18.8; Muris, Merckelbach et al. 2002). As commonly reported in the literature, females reported significantly higher levels of anxiety than males, *t* (158) = -3.44, *p* = .001, $M_{females}$ = 56.44, SD = 27.22, M_{males} = 42.08, SD = 25.28.

Life event measures Adolescents reported more positive than negative events on both the CHUS and BALES (Time 2a CHUS, t (228) = 21.85, p =.001, $M_{positive}$ = 38.25, SD = 13.39, $M_{negative}$ = 15.51, SD = 9.08. Time 2a BALES, t (217) = 21.89, p = .002, $M_{positive}$ = 10.94, SD = 4.91, $M_{negative}$ = 3.92, SD = 3.70[8]. This is in line with the development paper for the CHUS (Kanner, Feldman et al. 1987) but the reverse pattern from that observed in the BALES development paper (Shahar, Henrich et al. 2003). This difference in the relationship between positive and negative events on the BALES between the current study and original development paper seems to be driven by the lower frequency of reported negative events in the current sample compared to similar frequencies for positive events. Participants in the current sample also reported fewer negative events on the ALEQ (M = 12.56, SD = 8.56) compared to the original paper (M = 30.07, SD = 5.52; Hankin and Abramson 2002) despite the longer time frame captured in the current study (6 months compared to 5 weeks).

Internal consistency

All scales and subscales of the life event measures demonstrated adequate internal consistency with exception of the BALES Negative Achievement Events and ALEQ Interpersonal-independent subscales which approached the cut-off of .70 (Table 3). Values obtained in the current study were largely similar to those reported in the original

[8] Significantly positive than negative events were reported for the CHUS and BALES across all administrations (times 2a-2d). For simplicity, T tests reported for first administration (2a) only).

development papers. For example, alpha values were .87 for both hassles and uplifts subscales of the CHUS with 12-year olds (Kanner, Feldman et al. 1987) compared to .84 and .89, respectively, in the current study. Internal consistencies for the individual subscales of the ALEQ were not reported in the original sample of 13-19-year olds but the total scale estimate was .94, identical to that in the current study.

Test-retest reliability and repeated measurement

As recommended by the COSMIN checklist, intra-class correlations (ICCs) were used to examine the test-retest reliability of the BALES and the CHUS subscales at the first (Time 2a) and last (Time 2d) administration (Table 3). All subscales were lower than the recommended reliability coefficient of .70 (Terwee, Bot et al. 2007).

Change in the number of events endorsed in each subscale across the four weeks for both the BALES and CHUS is shown in Figure 2. For both measures, all subscale scores decreased significantly across measurements. The BALES demonstrated a significant reduction in Negative Interpersonal Events, $F(3,384) = 10.87$, $p < .01$; Negative Achievement Events, $F(3,384) = 13.07$, $p < .01$, Positive Interpersonal Events, $F(3,384) = 21.93$, $p < .01$ and Positive Achievement Events, $F(3,378) = 32.84$, $p < .01$. Similarly, the CHUS showed significant reductions in both Hassles, $F(3,405) = 30.78$, $p < .01$, and Uplifts, $F(3, 384) = 18.45$, $p < .01$, across the four weeks.

Table 3. Internal consistency and test-retest reliability of life event measures

		# items in subscale	Cronbach's Alpha (α)	Test-retest correlation (ICC)
BALES	Total		0.86	
	Negative Interpersonal Events	9	0.74	.51**
	Negative Achievement Events	9	0.69	.54**
	Positive Interpersonal Events	9	0.71	.36**
	Positive Achievement Events	9	0.84	.38**
CHUS	Total		0.87	
	Hassles	25	0.84	.51**
	Uplifts	25	0.89	.51**
ALEQ	Total		0.94	
	Interpersonal-dependent	27	0.89	--
	Interpersonal-independent	13	0.67	--
	Non-interpersonal	10	0.79	--

** $p < .01$

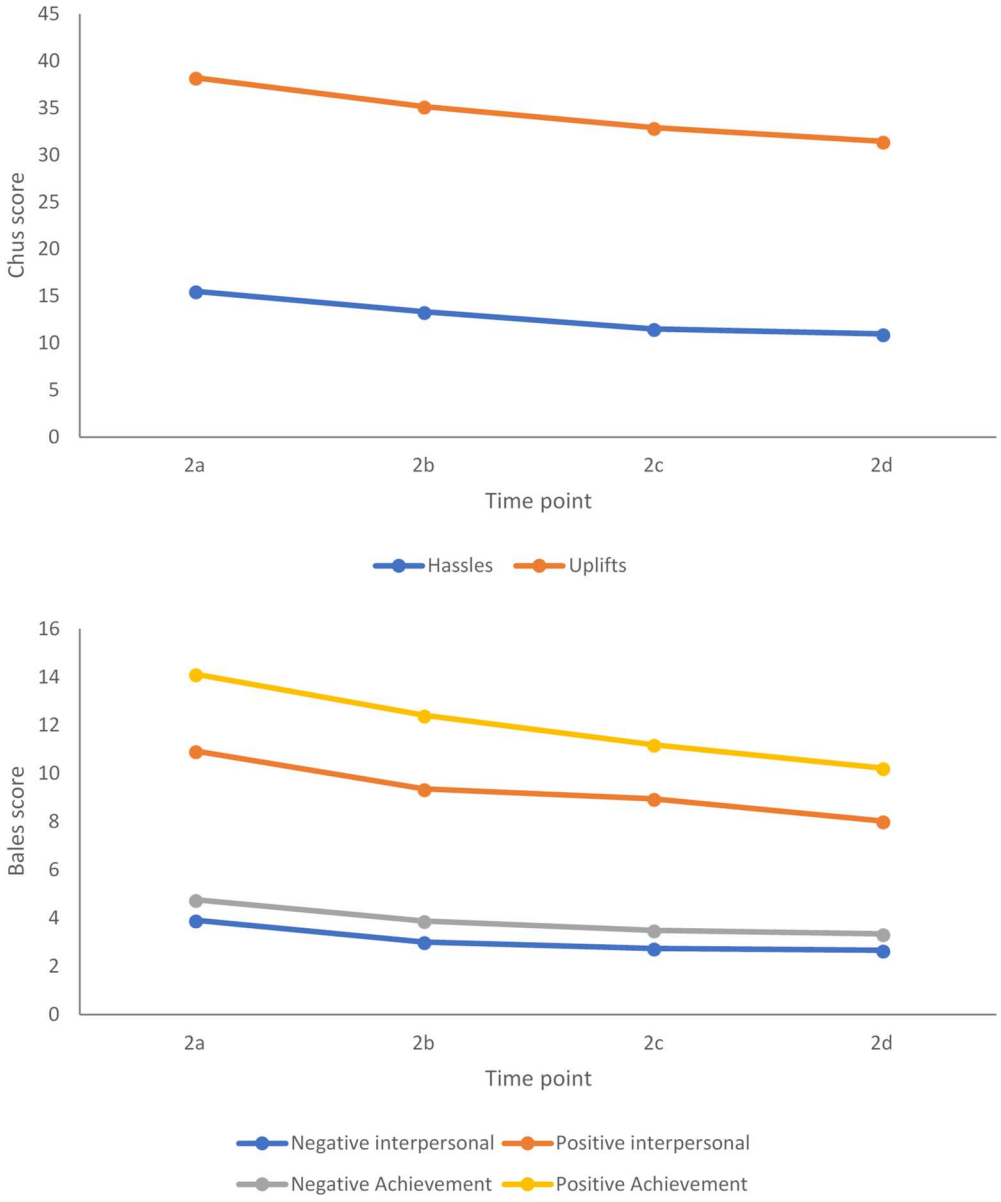


Figure 2. Daily hassles scores by week

Structural validity

Principal Components Analysis (PCA) was conducted for the first administration of each measure using varimax rotation, assuming uncorrelated factors.

CHUS (Kanner, Feldman et al. 1987)

The basic criteria for sampling adequacy (KMO = .73) and sphericity (Bartlett's χ^2 (1225) = 3334.37, $p < .001$) were satisfied. The scree plot identified two factors which explained 15.66% and 12.32% of the variance, respectively. In line with the originally proposed structure of the CHUS, these factors reflected positive (uplifts) and negative (hassles)

events, respectively (see Appendix 6). Most items had loadings greater than .3 on one of the two factors. However, two items had cross-loadings on both factors ("you had to clean your room", "there was a school trip") and two items did not clearly load on either factor ("you played with your pet", "you got a phone call").

BALES (Shahar, Henrich et al. 2003)

Estimates for sampling adequacy (KMO = .81) and sphericity (Bartlett's χ^2 (1225) = 3465.44, $p < .001$) were acceptable. Initial eigenvalues indicated that the first three factors explained 15.48%, 12.87% and 7.44% of the variance, respectively. However, closer inspection of the factors revealed little similarity to the originally proposed factor structure and an absence of semantic coherence (see Appendix 7). For example, whilst the first contained all items pertaining to positive events, it also had items from different life domains (e.g. "I got help from a family member when I needed it" and "I got a good grade at school"). Furthermore, the second factor contained both positive and negative events from across domains ("I made up with a friend" and "I got a bad grade at school"). A two-factor model, explaining 30.29% of the variance made greater semantic sense. Both factors had loadings greater than .30 for all items. As with the CHUS, factors generally reflected positive and negative events, respectively. However, three items pertaining to positive events (e.g. "I made up with a family member/friend") loaded more strongly on the negative event factors and one item ("I became sick or got injured") loaded more strongly on the positive event factor.

ALEQ (Hankin and Abramson 2002)

Sampling adequacy (KMO = .81) and sphericity (Bartlett's χ^2 (1596) = 6449.70, $p < .001$) were acceptable. Examination of the screen plot and initial eigenvalues revealed three factors explaining 25.34%, 7.48% and 5.17% of the variance, respectively. However, these did not reflect the subscales proposed by the original authors (i.e. interpersonal-independent, interpersonal-nonindependent, achievement events) and were not semantically coherent (see Appendix 8). For example, factor one included items related to both interpersonal (e.g. "Your girlfriend/ boyfriend criticized you") and achievement (e.g. "You got a bad report card or end of term report") events. PCA was repeated constraining items to two factors to see if separate subscales for dependent and independent events emerged. The total variance explained by the two-factor model was 32.82% (factors 1 and 2 explained 17.23% and 15.59% of the variance, respectively). However, these did not differentiate independent and dependent events as defined by the model and multiple items loaded on both factors.

Criterion/convergent validity

The COSMIN checklist defines criterion validity as the correlation of the instrument with a "gold standard" measure administered at the same time. However, since there is no 'gold standard' life event measure to compare the current life event measures to, we examined correlations between subscales on each measure to estimate the degree of convergence between them (Table 4). Williams' tests were used as recommended by Weaver and Wuensch (2013) to compare the relative magnitude of correlations between subscales.

Hypothesis 1 Daily hassles (CHUS, BALES) are more strongly positively correlated with each other than with major life events (ALEQ).

As predicted, the BALES and CHUS subscales were more strongly correlated with each other than either daily hassles measure with the ALEQ (e.g. BALES Negative Interpersonal with CHUS Hassles versus ALEQ Interpersonal-Dependent, t (163) = -5.99, p<.01; BALES Negative Achievement with CHUS Hassles versus ALEQ Non-Interpersonal, t(167) = -4.29, p<.01).

Hypothesis 2 Negative event subscales across measures will converge. Positive event subscales will converge.

As predicted, the BALES negative event subscales were more strongly correlated with CHUS hassles than uplifts subscales (e.g. BALES Negative Interpersonal with Hassles versus Uplifts, t (166) = -10.81, p< .01). The BALES positive subscales were both more strongly correlated with the CHUS uplifts subscale than the CHUS hassles subscale (e.g. BALES Positive Interpersonal with Hassles versus Uplifts, t (166) = -8.82, p < .01. Finally, the ALEQ subscales (which all pertain to negative events) were moderately correlated with the CHUS hassles but not uplifts subscales (e.g. ALEQ Interpersonal-Dependent with CHUS Hassles versus Uplifts, t(166) = -7.08, p < .01) and with the BALES negative but not positive subscales (e.g. ALEQ Interpersonal-Dependent with BALES Negative Interpersonal versus BALES Positive Interpersonal, t(165) = -5.76, p < .01).

Hypothesis 3 Interpersonal and non-interpersonal subscales will converge across measures.

As predicted, the BALES negative-interpersonal subscale correlated with both the ALEQ interpersonal subscales to a similar magnitude, t (166) = 1.03, *ns* and these were both significantly greater in magnitude than correlations with the ALEQ non-interpersonal subscale, t (166) = 2.49, p <.01. However, the ALEQ non-interpersonal scale was not more strongly correlated with the BALES negative achievement scale than the BALES negative interpersonal scale, t (166) =.76, *ns*.

Table 4. Intercorrelations (Pearson's r) between life event subscales

		1	2	3	4	5	6	7	8
1	BALES: Negative interpersonal events								
2	BALES: Negative achievement events	.766**							
3	BALES: Positive interpersonal events	.248**	.287**						
4	BALES: Positive achievement events	.122	.185*	.780**					
5	CHUS: hassles	.810**	.711**	.201**	0.086				
6	CHUS: uplifts	.106	0.135	.799**	.833**	0.082			
7	ALEQ: Interpersonal-dependent	.555**	.485**	0.068	-0.030	.555**	-0.063		
8	ALEQ: Interpersonal-independent	.499**	.493**	0.128	0.061	.483**	0.076	.635**	
9	ALEQ: non-interpersonal	.445**	.489**	0.122	-0.011	.521**	-0.051	.767**	.486**

** p <.01 ***p <.001. **BALES** Brief Adolescent Life Event Scale, **CHUS** Children's Hassles and Uplifts Scale, **ALEQ** Adolescent Life Event Questionnaire

Hypothesis testing

Pearson's correlations between life event measure subscales and symptoms of depression and anxiety at time one and depression at time three are shown in Table 5. Correlations were of similar magnitude to those reported in the original development papers (shown in grey in column 4). However, correlations were generally larger at the second administration of the CDI than the first (time 3, compared to time 1).

Table 5. Correlations (Pearson's r) between life event subscales and depression(CDI) and anxiety (MASC)

	Depression			Anxiety
	Time 1	Time 3		Time 1
			Shahar et al. (2003)	
BALES[1]				
Negative Interpersonal Experiences	.35**	.54**	.32**	.27**
Negative Achievement Events	.40**	.51**	.35**	.26**
Positive Interpersonal Events	-.10	-0.15	-0.12**	-.10
Positive Achievement Events	-.17 *	-.24**	-0.20**	-.16
CHUS[1]				
Hassles	.37**	.61**	--	.31**
Uplifts	-.22**	-.29**	--	-.12
			Hankin & Abramson (2002)	
ALEQ				
Interpersonal-dependent	.28**	.53**	.33***	.20*
Interpersonal-independent	.28**	.38**	.26***	.18*
Non-interpersonal	.29**	.42**	.37***	.29**

*p<.05, **p<.01 [1] mean across 4 time points **CDI** Children Depression Inventory; **MASC** Multidimensional Anxiety Scale for Children; **BALES** Brief Adolescent Life Event Scale; **CHUS** Children's Hassles and Uplifts Scale; **ALEQ** Adolescent Life Event Scale
Statistics in grey taken from original papers.

Hypothesis 1 Anxiety and depression symptoms will be positively correlated with negative event subscales and negatively correlated with positive event subscales.

As predicted, depression was positively correlated with all negative event subscales at time one and three (*rs* ranged between .28 - .54, *ps*<.01, Table 5, Column 2 & 3) and negatively correlated with positive event subscales (*rs* ranged between -.10 and .29), although the correlation between BALES positive interpersonal events and depression was not significant at either time one or three. Interestingly, positive events as measured with both the BALES and CHUS were not correlated with anxiety (*rs* ranged between -.10 and -.16, *ns*).

Hypothesis 2 Subscales pertaining to negative dependent, particularly interpersonal, events will be more strongly correlated with depression than independent or non-interpersonal subscales.

Contrary to our hypothesis, both the BALES and ALEQ subscales showed similar correlations between depression and both interpersonal and non-interpersonal subscales (e.g. BALES Negative Interpersonal versus BALES Negative Achievement at time 1; t (130) =-1.71, *ns*).

Hypothesis 3 Daily hassles subscales will be more strongly correlated with depression than major life events.

There were no significant differences in the magnitude of correlations with depression between daily hassles and major life event subscales (e.g. time 1 depression with BALES Negative Interpersonal versus ALEQ Interpersonal-Dependent events, t (127) = -.89 p<.01).

Discussion

Summary

The current study is the first quantitative comparison of the psychometric properties of multiple stressful life event measures in young people using the systematic COSMIN method. This represents an important development in research of stressful life events in young people for three main reasons. Firstly, it contributes to the evidence-base for each of the included measures by providing a detailed and systematic analysis of the full range of psychometric properties in a large population sample of adolescents. Secondly, it allows direct comparison between the three measures with regards to their relative psychometric strengths to aid researchers in selecting between them in future research. Thirdly and finally, the current study highlights appropriate methods for systematically assessing psychometric properties of stressful life event measures in young people. Results are summarised in Table 6. More detailed discussion of each psychometric property follows together with recommendations for future research.

Table 6. Summary of findings for psychometric properties of stressful life event measures

Psychometric property	Criteria/hypotheses	CHUS	BALES	ALEQ
Content validity	"bottom-up" development (using young people to generate items)	✓ Items generated from semi-structured interviews with young people	✗ Items generated by researchers	✗ Not reported
Internal consistency	Cronbach's alpha >.70	✓	? Some subscales <.70	✓
Test-retest reliability	ICC >.70	✗	✗	–
Structural validity	PCA identifies semantically coherent subscales	✓ Two factors representing 'hassles' and 'uplifts'	? PCA identified negative and positive factors but not different life domains	✗ No semantically meaningful subscales identified
Criterion/Convergent validity	Negative vs. positive events	✓ Convergence with negative/positive subscales across CHUS and BALES		–
	Hassles vs major life events	✓ Greater convergence between subscales on CHUS and BALES than with ALEQ subscales		
	Subscale specificity	–	? ALEQ interpersonal subscales converged with BALES interpersonal but ALEQ non-interpersonal similar correlations with BALES interpersonal and non-interpersonal subscales	
Hypothesis testing	Negative vs positive events	✓	✓	✓ (doesn't include positive events)
	Hassles vs major life events	✗ Magnitude of correlations similar for hassles and major life events		
	Subscale specificity	–	✗ Interpersonal and non-interpersonal similar magnitude correlations	✗ Interpersonal and non-interpersonal similar magnitude correlations

✓ met criteria ✗ did not meet criteria ? partially met criteria – not possible to assess

Content validity

Content validity refers to the degree to which the items reflect the construct to be measured in the target population. It was not possible to formally assess content validity in the current study. However, the content validity for each measure is discussed here with reference to the initial development paper for each measure and broader conceptual issues outlined in the introduction.

There are two main criteria for assessing content validity; relevance and comprehensiveness. Relevance considers whether all included items reflect aspects of the construct to be measured and are relevant to the target population. Comprehensiveness refers to whether all items comprehensively reflect the construct to be measured.

For relevance, the COSMIN taxonomy recommends using 'bottom-up' scale development methods in which people from the target population are used to generate a pool of possible items and that an independent sample from the same population should rate the relevance of included items. Of the three measures examined, only one, the CHUS measure of daily events, met the COSMIN criterion of 'bottom-up' item generation. Items were generated from responses of children and early adolescents to a semi-structured interview about stress in their lives, and then factor analysis was used to assess whether hassle and uplift items load on separate factors (Kanner, Feldman et al. 1987). The items on the BALES were originally derived by the authors from existing measures (Shahar, Henrich et al. 2003) whilst the method of measure development for the ALEQ is not described (Hamilton, Stange et al. 2013). Therefore, it is possible that the BALES and ALEQ contain items that are not relevant in the lives of young people in the target age range and this would negatively impact psychometric properties such as internal consistency and structural validity. For example, item-level scrutiny of the current measures highlighted that some items may not be suitable for the intended age group. For example, items such as "you had a romantic date" on the BALES and "You Did something to please your boyfriend/ girlfriend that you didn't want to do" on the ALEQ are unlikely to be relevant for the youngest participants and indeed some raised this with researchers during data collection.

Examining the comprehensiveness of the three studied measures of stress, it is a challenge for any one measure to cover all domains and aspects of stress, given the breadth and multi-dimensional nature of the construct. Assessment of the comprehensiveness of a measure of stressful life events will depend on the conceptual definition used. For example,

measures included in the current study were designed to capture either daily hassles or major life events. The CHUS and BALES, both designed to measure daily hassles, seem to capture more everyday irritating/distressing events (e.g. CHUS #8 “you lost something” and BALES #13 “a classmate teased me”). However, the ALEQ captures some major events (e.g. #3“a close family member died) but also items that overlap with those covered by BALES and CHUS, reflecting more daily hassles (e.g. #11 “you didn’t spend as much time with a close family member as you wanted to”). The diversity of possible life domains (e.g. family, peer, school, appearance etc.) results in an almost infinite number of possible of events/experiences that could be measured. This has resulted in researchers covering different domains to varying extents. For example, the BALES included a subset of items relating to appearance concerns which do not appear in the CHUS or ALEQ.

The relevance and comprehensiveness of stressful life event measures is especially important given evidence for specificity in relationships between certain aspects of stressful life events and psychopathology in young people. Future research would benefit from a more detailed understanding of the relevance of specific stressful life events in different populations of youth. To do this, it would be useful to ask samples of young people to rate the relevance of the items included across multiple measures.

Alternatively, given the challenge of creating comprehensive measures of stressful life events due to the broad and multifaceted nature of the concept, it could be helpful if measures of stressful life events had a tight focus and scope, so rather than a measure of stressful life events *per se*, they were designed to capture major life events, or interpersonal events, or hassles specifically.

Structural validity and Internal consistency

Assessment of structural validity is related to content validity; whether the subscales defined within the measures represent the underlying structure of the target conceptualisation of stressful life events. Factor analysis using PCA generally distinguished positive and negative events on the CHUS and BALES but did not differentiate more specific, semantically coherent scales like those proposed by the original authors. However, Cronbach’s alpha estimates for the originally proposed subscales indicated adequate internal consistency in the current study, suggesting that items within the subscales are statistically more related than would be expected by chance.

There are theoretical issues with using statistical analyses to determine the structural validity and internal consistency of life event measures. First, both analyses assume

reflective models in which items correlate because they co-occur. However, whilst life events may be semantically linked (e.g. reflecting the same life domain) they do not necessarily co-occur (Compas, 1987). For example, 'arguing with a friend' does not necessarily make you more likely to have 'argued with a family member' despite both items appearing in the BALES negative-interpersonal experiences subscale. In these cases, inter-item correlations could be near-zero, and so would not cluster together in factor analyses and would serve to decrease estimates of internal consistency.

Another issue for the assessment of structural validity and internal consistency of life event measures is that the actual experiences leading to respondents endorsing an item on a checklist vary greatly between individuals, so called intra-categorical variability (Dohrenwend, 2006). It is possible that two individuals endorsing the same life event may have had different experiences relating to that event. For example, one study interviewed adults after they completed a life event checklist to identify the experiences relating to the events they endorsed (Dohrenwend, 2006). They found that reasons for endorsing items varied greatly between individuals, for example experiences relating to the item "serious illness or injury" ranged from episodes of flu to severe heart attacks.

This is important because the actual experiences behind endorsed checklist items may vary in meaningful ways that confound their associations with psychological outcomes (Dohrenwend, 2006). Experiences may vary in their severity (i.e. minor versus major life events) and valence (i.e. pleasantness-aversiveness) as well as in the level of dependence on the individual (i.e. whether the event was due to environmental circumstances or the actions of the individual). This is particularly relevant for stress-depression research which proposes that dependent life events have greater influence on depression than independent events.

Given difficulties with the statistical assessment of the psychometric properties, it may be more important to focus on semantic coherence of stressful life event measures. As a result, it is especially important to establish content validity prior to any further examination of psychometric properties and utility. This is rarely done in stressful life events research with researchers often creating or adapting existing measures for their specific study aims without formally assessing their comprehensiveness and their relevance to the target population. This could be achieved with the greater involvement of young people to generate the items and rate comprehensiveness and relevance to constructs being measured.

Test-retest reliability

As recommended, ICCs were calculated instead of Pearson's coefficients to account for systematic error (Terwee, Bot et al. 2007). Test-retest correlations for all subscales were lower than the recommended cut-off of .70 indicating a lack of consistency in participants' responding over time. Consistent responding over the time frame of few weeks might not be expected for daily hassles which are typically defined as more discrete everyday events. As a result, they may not persist from one week to the next.

However, instability in reports of stressful life events across the CHUS and BALES was also highlighted in reductions in the mean number of events reported across the four weekly measurements. Reductions in mean scores over time could indicate testing or measurement error. Multiple sources of error have been suggested including social desirability, decreased test anxiety and habituation or boredom with completing the measure (Twenge and Nolen-Hoeksema 2002). Only one other study, to our knowledge, has examined effects of repeated measurements of life events. Wittchen and colleagues (Wittchen, Essau et al. 1989) found significant reductions in the average number of stressful life events reported by adults on the Munich Event List over 6 years. More specifically, they found greater instability for daily hassles compared to major life events and for positive compared to negative events. However, both negative and positive events on the CHUS and BALES showed similar reductions across time points in the current study. It was not possible to compare relative stability of daily hassles and major life events in the current study because the ALEQ was only administered on one occasion. It is possible that reduced means across repeated measurements for the CHUS and BALES in the current study could reflect fatigue/boredom since participants were asked to complete multiple life events measures on several occasions.

There is also debate around the appropriate methods for assessing the reliability of life event measures, with some researchers opting for correlations between total (sub)scale scores over time whilst others examine item-level percentage agreement or use inter-rater reliability by, for example, examining agreement between child and either parent or teacher reports (Glen, Simpson et al. 1993). However, comparing multiple informants of young people's life events is problematic as other informants may not observe all life domains where events are experienced, especially in adolescence where individuals spend a greater proportion of time away from adult caregivers.

In the absence of gold standard methods for assessing the reliability of stressful life event measures, it is advisable that future research include multiple analyses of reliability such as test-retest correlations, comparisons of means across measurement and item-level agreement across time. Furthermore, given evidence for drop off in the numbers of events reported across measurements in the current study, it will be important for future longitudinal research to measure and report changes in stressful life event scores across administrations.

Criterion Validity

Given the lack of a 'gold standard' stressful life event measure in adolescence, we examined the degree of convergence between the three selected measures. In the current study, as expected, daily hassles subscales more strongly correlated with each other than with the ALEQ subscales. Both daily hassles measures contained similar items and were designed to capture the same time frame of one week, whilst the ALEQ captured more major life events spanning the previous six months. However, the CHUS and BALES were also both completed at time two whilst the ALEQ was completed at time three, only one week following completion of time two. This could artificially deflate the degree of convergence between responses on the daily hassles measures and the ALEQ.

As expected, subscales pertaining to negative events across measures showed moderate to large correlations with each other and subscales pertaining to positive events were also moderately correlated with one another whilst negative and positive event subscales showed near-zero correlations with each other, suggesting that the presence of negative events was not associated with the presence of positive events.

There was a lack of specificity in associations between subscales depicting similar life domains such as interpersonal subscales on the BALES and ALEQ. This may reflect problems with intra-categorical variability where the actual event leading to a respondent endorsing the item may vary in meaningful ways (Dohrenwend 2006). For example, for the item "a classmate teased me", although this clearly reflects an interpersonal life event, it could be categorised as either independent, if assumed unprovoked, or dependent, if the individual was involved in conflict with the classmate. This would result in different response patterns between individuals which would decrease specificity in associations between dependent-independent subscales.

Hypothesis Testing

As predicted, negative event subscales on all three measures were moderately positive correlated with both depression and anxiety whilst positive event subscales were negatively correlated with depression, but not anxiety. This is consistent with studies of university students which have consistently shown associations between anxiety and negative, but not positive, events (Sarason, Johnson et al. 1978, Nezu 1986). Reward models of depression propose that acute and chronic stress (i.e. negative life events) contributes to depression but also a reduced frequency of rewarding/pleasurable (i.e. positive) events (Auerbach, Admon et al. 2014). Furthermore, the experience of positive events is proposed to have a buffering effect on depression (Cohen, McGowan et al. 1984, Dixon and Reid 2000). In contrast, anxiety may not be associated with a general reduction in the occurrence of positive events and instead, more associated with avoidance of specific negative situations. However, research into the role of positive events is limited and Kashdan & Sieger (2006) found that greater social anxiety was associated with fewer daily positive events in their experience sampling study of university students.

Contrary to previous research, the current study found that depression was similarly associated with both daily hassles and major life events and with both independent and non-independent events. This suggests a lack of specificity in the associations between depression and different types of life events in adolescents.

Combined, the results obtained for assessment of criterion validity and hypothesis testing showed that the current stressful life event measures demonstrate (i) differential associations for positive and negative event subscales and (ii) greater convergence between daily hassles measures when compared to a measure of major life events. However, the measures are not able to detect specificity in relationships between different domains of stressful life events (e.g. dependent-independent, interpersonal-non-interpersonal) and symptom measures. Therefore, it remains unclear whether this reflects the true nature of relationships between stressful life events and symptoms of depression and anxiety or measurement issues.

Strengths and Limitations

The large sample size, inclusion of multiple life event measures within a single sample, and the systematic evaluation of the psychometric properties using best-practice guidelines are all considerable strengths of the current study. However, the study also had some limitations.

The broader research project aimed to examine stress-vulnerability hypotheses of depression in adolescents and so, was not specifically designed to examine psychometric properties of stressful life events. The design constraints imposed by this broader research aim limited the evaluation of some psychometric properties in the current study. For example, because the ALEQ was only administered once we could not determine the test-retest reliability of the scale and we were not able to examine content validity as recommended by COSMIN by asking focus groups of young people to assess relevance and comprehensiveness of the life event measures.

Furthermore, both the CHUS and BALES were administered four times with the initial aim to allow more detailed examination of the accumulation of daily hassles over time as part of the wider stress-vulnerability research project. However, repeated administration of the same measure may have resulted in participants becoming bored with completing it and therefore invalidating the responses given. Both the CHUS and BALES demonstrated reduced endorsement for both positive and negative items across the four administrations. This may negatively affect assessment of psychometrics in the current study, particularly estimates of test-retest reliability which compared responses on the first and fourth administration of each measures.

Using an opt-out consent procedure allowed us to maximise the sample across the six waves of data collection and ensure a representative sample without systematic non-participation. However, participation at each time point was dependent on how many young people were present in class on the days of data collection and how many of these provided verbal assent at each time point. This resulted in a complex pattern of missing data across the six time points, with a minority of young people completing some but not all of the measures included in the current study. However, comparison of demographic (age, gender, ethnicity) and symptoms measures (anxiety and depression) demonstrated no significant differences between those who completed time one but not time two and/or time three, suggesting that the sample was representative of the wider school population at each stage of data collection.

Finally, minor changes were made to the stressful life event measures. For example, the response scale on CHUS was changed from the original measure of impact (i.e. how good/bad the event was) to match the BALES, measuring how frequently the event occurred. Although this impedes direct comparison with previous studies using the CHUS, it facilitated direct comparison between the performance of the CHUS and BALES in a single

sample by removing the possible confounding effects of difference in response scale. This is especially important given differences between subjective and objective conceptualisations of stress. Frequency ratings represent more objective measurement of life events compared to impact ratings which involve subjective appraisal of the valence of the event as positive or negative.

Recommendations

Aim 1: Facilitate the selection of life event measures

The main aim of the current study was to aid researchers with the selection of life event measures for use with young people. Of the three measures, the CHUS demonstrated the most favourable psychometric properties (Table 6). More specifically, the CHUS was the only measure with adequate content validity. The CHUS also showed adequate structural validity and internal consistency. However, as with all the measures, test-retest reliability estimates were inadequate. There was evidence of convergent validity with the BALES subscales and of hypothesised relationships with symptom measures of anxiety and depression. Conversely, the ALEQ demonstrated more limited evidence for adequate psychometrics. The content validity of the ALEQ is questionable because the method of scale development has not been reported and the relevance of items has not been assessed in the target population. Closer comparison of items across the three measures indicated that the ALEQ may include measures of both major life events and daily hassles. Furthermore, using factor analysis, there was no evidence of structural validity for the proposed subscales and these subscales did not show specificity in associations with depression and anxiety as hypothesised.

Whilst these results might suggest that researchers should avoid using the ALEQ and instead opt for the CHUS, other factors beyond psychometric properties, such as the specific conceptualisation of stress, should be considered when selecting an appropriate measure. Stress is multi-dimensional, and it is unlikely that one measure can adequately capture all facets. Future research would benefit from the development of a clear working definition of stressful life events. Grant and colleagues (2003) propose a systematic process for developing a taxonomy of stressful life events. First, interviews should be conducted with young people to identify objective stressful life events with high interrater agreement. Second, responses on these interviews should be used to create a list of stressful life events. These responses should then be trialled in different samples of young people to assess their comprehensiveness and relevance to various subsamples (e.g. young children, children

living in different environmental contexts). Finally, the measures should be administered to large samples to derive population norms from which to compare subsequent studies.

Alternatively, developing a comprehensive and agreed taxonomy of the various conceptualisations of stressful life events would allow the development of more specific life event measures designed to capture particular facets of life events, for example measures specific to daily hassles or major life events within specific life domains such as interpersonal or achievement related stressors.

Aim 2: psychometric analyses for life event measures

The second aim of the current study was to highlight appropriate methods for the assessment of psychometric properties of stressful life event measures. Several recommendations can be made.

More thorough examination of psychometric properties is a required standard for future stressful life events research. Basic research assessing the psychometric properties and performance of different life event measures in multiple samples of children and adolescents using standardised methods would enable increased confidence in research findings and clarify our understanding of the complex role of life events in developmental psychopathology. Whilst the use of the COSMIN checklist represents a significant step forward in systematising psychometric assessment, there are theoretical issues in applying this taxonomy directly to life event measures. For example, the questionable value of using statistical assessment of structural validity and internal consistency with discrete life event items.

Future research would benefit from greater theoretical consideration of which psychometrics properties are important and appropriate to examine. This would allow the development of clear best-practice standards for judging stressful life event measures. Given the lack of consensus for the best methods for determining the psychometric properties of stressful life event measures, it might be useful to compare different analyses on the same measure (e.g. examining test-retest reliability using correlations between time points as well as item-level agreement). It might also be useful to compare multiple formats of life events measures (e.g. comparing checklists with interviews) and examine responses across multiple informants (e.g. child, parent, teacher).

Conclusions

The current study adds to the body of knowledge regarding the role of stressful life events in the development of psychological problems in young people. The systematic evaluation of the psychometric properties of three established life event measures in a large population sample of adolescents using the COSMIN standards highlighted theoretical issues with the assessment of psychometric properties of life event measures. It is hoped that the findings of this study and the recommendations made will help to reduce inconsistency in the field and ultimately enhance our understanding of the complex interplay between life events and psychopathology in young people.

References

Allen, J. L., Rapee, R. M., & Sandberg, S. (2012). Assessment of maternally reported life events in children and adolescents: a comparison of interview and checklist methods. *Journal of psychopathology and behavioral assessment, 34*(2), 204-215.

Auerbach, R. P., Admon, R., & Pizzagalli, D. A. (2014). Adolescent depression: stress and reward dysfunction. *Harvard review of psychiatry, 22*(3), 139.

Bagley, C. (1992). Development of an Adolescent Stress Scale for Use by School Counsellors Construct Validity in Terms of Depression, Self-Esteem and Suicidal Ideation. *School Psychology International, 13*(1), 31-49.

Bailey, D., & Garralda, M. E. (1990). Life events: Children's reports. *Social psychiatry and psychiatric epidemiology, 25*(6), 283-288.

Bentler, P. M. (2007). On tests and indices for evaluating structural models. *Personality and Individual Differences, 42*(5), 825-829.

Birmaher, Sermiento, Ryan, Perel, Sanabria et al. (1996). Childhood and adolescent depression: A review of the past 10 years. Part I. *Journal of the Amerian Academy of Child & Adolescent Psychiatry, 35(11),* 1427-1439.

Blount, R. L., Simons, L. E., Devine, K. A., Jaaniste, T., Cohen, L. L., Chambers, C. T., & Hayutin, L. G. (2007). Evidence-based assessment of coping and stress in pediatric psychology. *Journal of pediatric psychology, 33*(9), 1021-1045.

Bobo, J. K., Gilchrist, L. D., Elmer, J. F., Snow, W. H., & Schinke, S. P. (1986). Hassles, role strain, and peer relations in young adolescents. *The Journal of Early Adolescence, 6*(4), 339-352.

Brown, H., Smith, P., & Leigh, E. (in prep). A systematic review of measures of stressful life events in young people.

Byrne, D. G., Davenport, S., & Mazanov, J. (2007). Profiles of adolescent stress: The development of the adolescent stress questionnaire (ASQ). *Journal of adolescence, 30*(3), 393-416.

Chance, M. (1982). CopingBehaviors That Mediate Between Life Events and Depression. *Arch Gen Psychiatry, 39*, 1386-1391.

Cheng, C. (1997). Assessment of major life events for Hong Kong adolescents: The Chinese adolescent life event scale. *American Journal of Community Psychology, 25*(1), 17-33.

Coddington, R. D. (1972). The significance of life events as etiologic factors in the diseases of children: I—A survey of professional workers. *Journal of Psychosomatic Research, 16*(1), 7-18.

Cohen, L. H., McGowan, J., Fooskas, S., & Rose, S. (1984). Positive life events and social support and the relationship between life stress and psychological disorder. *American Journal of Community Psychology, 12*(5), 567-587.

Cohen, S., Kessler, R. C., & Gordon, L. U. (1997). *Measuring stress: A guide for health and social scientists*: Oxford University Press on Demand.

Colton, J. A. (1985). Childhood stress: Perceptions of children and professionals. *Journal of psychopathology and behavioral assessment, 7*(2), 155-173.

Compas, B. E. (1987). Stress and life events during childhood and adolescence. *Clinical Psychology Review, 7*(3), 275-302.

Compas, B. E., Connor-Smith, J., & Jaser, S. S. (2004). Temperament, stress reactivity, and coping: Implications for depression in childhood and adolescence. *Journal of Clinical Child and Adolescent Psychology, 33*(1), 21-31.

Compas, B. E., Davis, G. E., Forsythe, C. J., & Wagner, B. M. (1987). Assessment of major and daily stressful events during adolescence: the Adolescent Perceived Events Scale. *Journal of consulting and clinical psychology, 55*(4), 534.

Copeland, W. E., Keeler, G., Angold, A., & Costello, E. J. (2007). Traumatic events and posttraumatic stress in childhood. *Archives of General Psychiatry, 64*(5), 577-584.

Dixon, W. A., & Reid, J. K. (2000). Positive Life Events as a Moderator of Stress-Related Depressive Symptoms. *Journal of Counseling & Development, 78*(3), 343-347.

Dodge, K. A., & Pettit, G. S. (2003). A biopsychosocial model of the development of chronic conduct problems in adolescence. *Developmental psychology, 39*(2), 349.

Dohrenwend, B. P. (2006). Inventorying stressful life events as risk factors for psychopathology: Toward resolution of the problem of intracategory variability. *Psychological bulletin, 132*(3), 477.

Duggal, S., Malkoff-Schwartz, S., Birmaher, B., Anderson, B. P., Matty, M. K., Houck, P. R., . . . Frank, E. (2000). Assessment of life stress in adolescents: Self-report versus interview methods. *Journal of the American Academy of Child & Adolescent Psychiatry, 39*(4), 445-452.

Eley, T. C., & Stevenson, J. (2000). Specific life events and chronic experiences differentially associated with depression and anxiety in young twins. *Journal of abnormal child psychology, 28*(4), 383-394.

Field, A. (2013). *Discovering statistics using IBM SPSS statistics*: Sage Publications, Ltd: UK.

Finlay-Jones, R., & Brown, G. W. (1981). Types of stressful life event and the onset of anxiety and depressive disorders. *Psychological Medicine, 11*(4), 803-815.

Ge, X., Conger, R. D., & Elder Jr, G. H. (2001). Pubertal transition, stressful life events, and the emergence of gender differences in adolescent depressive symptoms. *Developmental psychology, 37*(3), 404.

Ge, X., Lorenz, F. O., Conger, R. D., Elder, G. H., & Simons, R. L. (1994). Trajectories of stressful life events and depressive symptoms during adolescence. *Developmental psychology, 30*(4), 467.

Glen, S., Simpson, A., Drinnan, D., McGuinness, D., & Sandberg, S. (1993). Testing the reliability of a new measure of life events and experiences in childhood: The Psychosocial Assessment of Childhood Experiences (PACE). *European Child & Adolescent Psychiatry, 2*(2), 98-110.

Grant, K. E., Compas, B. E., Stuhlmacher, A. F., Thurm, A. E., McMahon, S. D., & Halpert, J. A. (2003). Stressors and child and adolescent psychopathology: Moving from markers to mechanisms of risk. *Psychological bulletin, 129*(3), 447.

Grant, K. E., Compas, B. E., Thurm, A. E., McMahon, S. D., & Gipson, P. Y. (2004). Stressors and child and adolescent psychopathology: Measurement issues and prospective effects. *Journal of Clinical Child and Adolescent Psychology, 33*(2), 412-425.

Hamilton, J. L., Stange, J. P., Shapero, B. G., Connolly, S. L., Abramson, L. Y., & Alloy, L. B. (2013). Cognitive vulnerabilities as predictors of stress generation in early adolescence: pathway to depressive symptoms. *Journal of abnormal child psychology, 41*(7), 1027-1039.

Hammen, C., & Rudolph, K. D. (1996). Childhood depression. In J. Mash, Barkley, R. A. (Ed.), *Child Psychopathology*. New York: Guildford.

Hankin, B. L., & Abramson, L. Y. (2002). Measuring cognitive vulnerability to depression in adolescence: reliability, validity, and gender differences. *J Clin Child Adolesc Psychol, 31*(4), 491-504.

Hankin, B. L., Mermelstein, R., & Roesch, L. (2007). Sex differences in adolescent depression: Stress exposure and reactivity models. *Child development, 78*(1), 279-295.

Harrington, D. (2008). Confirmatory factor analysis: Oxford university press. In: USA.

Holroyd, J. (1974). The Questionnaire on Resources and Stress: An instrument to measure family response to a handicapped family member. *Journal of community psychology, 2*(1), 92-94.

Hovens, J. G., Wiersma, J. E., Giltay, E. J., Van Oppen, P., Spinhoven, P., Penninx, B. W., & Zitman, F. G. (2010). Childhood life events and childhood trauma in adult patients with depressive, anxiety and comorbid disorders vs. controls. *Acta Psychiatrica Scandinavica, 122*(1), 66-74.

Hu, L. t., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural equation modeling: a multidisciplinary journal, 6*(1), 1-55.

IBM Corp. (2016). SPSS for Windows (Version 24.0). Armonk, NY: IBM Corp.

Johnson, J. H., & McCutcheon, S. M. (1980). Assessing life stress in older children and adolescents: preliminary findings with the life events checklist. In S. I. G. & S. C. D. (Eds.), *Stress and Anxiety*. Washington DC: Hemisphere.

Johnston, C. A., Steele, R. G., Herrera, E. A., & Phipps, S. (2003). Parent and child reporting of negative life events: discrepancy and agreement across pediatric samples. *Journal of Pediatric Psychology, 28*(8), 579-588.

Kanner, A. D., Coyne, J. C., Schaefer, C., & Lazarus, R. S. (1981). Comparison of two modes of stress measurement: Daily hassles and uplifts versus major life events. *Journal of behavioral medicine, 4*(1), 1-39.

Kanner, A. D., Feldman, S. S., Weinberger, D. A., & Ford, M. E. (1987). Uplifts, hassles, and adaptational outcomes in early adolescents. *The Journal of Early Adolescence, 7*(4), 371-394.

Kashdan, T. B., Barrios, V., Forsyth, J. P., & Steger, M. F. (2006). Experiential avoidance as a generalized psychological vulnerability: Comparisons with coping and emotion regulation strategies. *Behaviour research and therapy, 44*(9), 1301-1320.

Kearney, C. A., Drabman, R. S., & Beasley, J. F. (1993). The trials of childhood: The development, reliability, and validity of the Daily Life Stressors Scale. *Journal of Child and Family Studies, 2*(4), 371-388.

Kercher, A. J., Rapee, R. M., & Schniering, C. A. (2009). Neuroticism, life events and negative thoughts in the development of depression in adolescent girls. *Journal of abnormal child psychology, 37*(7), 903-915.

Kersten, P., Czuba, K., McPherson, K., Dudley, M., Elder, H., Tauroa, R., & Vandal, A. (2016). A systematic review of evidence for the psychometric properties of the Strengths and Difficulties Questionnaire. *International Journal of Behavioral Development, 40*(1), 64-75.

Kim, K. J., Conger, R. D., Elder Jr, G. H., & Lorenz, F. O. (2003). Reciprocal influences between stressful life events and adolescent internalizing and externalizing problems. *Child development, 74*(1), 127-143.

Kovacs, M. (1982). *Children's depression inventory—short version*: North Tonawanda, NY: Multi-Health Systems Inc.

Kushner, S. C., & Tackett, J. L. (2017). Parent and adolescent agreement for reports of life stressors. *Assessment, 24*(2), 143-156.

Lazarus, R. S. (1990). Theory-based stress measurement. *Psychological inquiry, 1*(1), 3-13.

Lewinsohn, P. M., & Talkington, J. (1979). Studies on the measurement of unpleasant events and relations with depression. *Applied Psychological Measurement, 3*(1), 83-101.

March, J. S., Parker, J. D., Sullivan, K., Stallings, P., & Conners, C. K. (1997). The Multidimensional Anxiety Scale for Children (MASC): factor structure, reliability, and validity. *Journal of the American academy of child & adolescent psychiatry, 36*(4), 554-565.

Miller, D. B., & Townsend, A. (2005). Urban hassles as chronic stressors and adolescent mental health: The Urban Hassles Index. *Brief Treatment and Crisis Intervention, 5*(1), 85.

Miller, D. B., Webster, S. E., & MacIntosh, R. (2002). What's there and what's not: Measuring daily hassles in urban African American adolescents. *Research on Social Work Practice, 12*(3), 375-388.

Moher, D., Shamseer, L., Clarke, M., Ghersi, D., Liberati, A., Petticrew, M., . . . Stewart, L. A. (2015). Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015 statement. *Systematic Reviews, 4*(1), 1.

Mokkink, L. B., Prinsen, C. A., Bouter, L. M., de Vet, H. C., & Terwee, C. B. (2016). The COnsensus-based Standards for the selection of health Measurement INstruments (COSMIN) and how to select an outcome measurement instrument. *Brazilian journal of physical therapy*(AHEAD), 0-0.

Mokkink, L. B., Prinsen, C. A., Bouter, L. M., Vet, H. C., & Terwee, C. B. (2016). The COnsensus-based Standards for the selection of health Measurement INstruments (COSMIN) and how to select an outcome measurement instrument. *Brazilian Journal of Physical Therapy, 20*(2), 105-113.

Mokkink, L. B., Terwee, C. B., Patrick, D. L., Alonso, J., Stratford, P. W., Knol, D. L., . . . De Vet, H. C. (2010a). The COSMIN checklist for assessing the methodological quality of studies on measurement properties of health status measurement instruments: an international Delphi study. *Quality of life research, 19*(4), 539-549.

Mokkink, L. B., Terwee, C. B., Patrick, D. L., Alonso, J., Stratford, P. W., Knol, D. L., . . . de Vet, H. C. (2010b). The COSMIN study reached international consensus on taxonomy, terminology, and definitions of measurement properties for health-related patient-reported outcomes. *Journal of clinical epidemiology, 63*(7), 737-745.

Moksnes, U. K., Byrne, D. G., Mazanov, J., & Espnes, G. A. (2010). Adolescent stress: evaluation of the factor structure of the adolescent stress questionnaire (ASQ-N). *Scandinavian Journal of Psychology, 51*(3), 203-209.

Monck, E., & Dobbs, R. (1985). Measuring life events in an adolescent population: methodological issues and related findings. *Psychol Med, 15*(4), 841-850.

Muris, P., Merckelbach, H., Ollendick, T., King, N., & Bogie, N. (2002). Three traditional and three new childhood anxiety questionnaires: Their reliability and validity in a normal adolescent sample. *Behaviour research and therapy, 40*(7), 753-772.

Nezu, A. M. (1986). Negative life stress and anxiety: Problem solving as a moderator variable. *Psychological reports, 58*(1), 279-283.

Nilsson, D., Gustafsson, P. E., Larsson, J. L., & Svedin, C. G. (2010). Evaluation of the Linköping youth life experience scale. *The Journal of nervous and mental disease, 198*(10), 768-774.

Pagano, M. E., Skodol, A. E., Stout, R. L., Shea, M. T., Yen, S., Grilo, C. M., . . . Zanarini, M. C. (2004). Stressful life events as predictors of functioning: findings from the Collaborative Longitudinal Personality Disorders Study. *Acta Psychiatrica Scandinavica, 110*(6), 421-429.

Platt, R., Williams, S. R., & Ginsburg, G. S. (2016). Stressful Life Events and Child Anxiety: Examining Parent and Child Mediators. *Child Psychiatry and Humum Development, 47*(1), 23-34.

Rende, R. D., & Plomin, R. (1991). Child and parent perceptions of the upsettingness of major life events. *Journal of Child Psychology and Psychiatry, 32*(4), 627-633.

Rudolph, K. D., & Hammen, C. (1999). Age and gender as determinants of stress exposure, generation, and reactions in youngsters: A transactional perspective. *Child development, 70*(3), 660-677.

Rudolph, K. D., Hammen, C., Burge, D., Lindberg, N., Herzberg, D., & Daley, S. E. (2000). Toward an interpersonal life-stress model of depression: The developmental context of stress generation. *Development and psychopathology, 12*(2), 215-234.

Rutter, M. (1981). Stress, coping and development: Some issues and some questions. *Journal of child psychology and psychiatry, 22*(4), 323-356.

Sarason, I. G., Johnson, J. H., & Siegel, J. M. (1978). Assessing the impact of life changes: development of the Life Experiences Survey. *Journal of consulting and clinical psychology, 46*(5), 932.

Seiffge-Krenke, I. (2000). Causal links between stressful events, coping style, and adolescent symptomatology. *Journal of adolescence, 23*(6), 675-691.

Shahar, G., Henrich, C. C., Reiner, I. C., & Little, T. D. (2003). Development and initial validation of the brief adolescent life event scale (BALES). *Anxiety, Stress & Coping, 16*(1), 119-128.

Stikkelbroek, Y., Bodden, D. H. M., Kleinjan, M., Reijnders, M., & van Baar, A. L. (2016). Adolescent Depression and Negative Life Events, the Mediating Role of Cognitive Emotion Regulation. *PLoS One, 11*(8).

Streiner, D. L. (2003). Starting at the beginning: an introduction to coefficient alpha and internal consistency. *Journal of personality assessment, 80*(1), 99-103.

Terwee, C. B., Bot, S. D., de Boer, M. R., van der Windt, D. A., Knol, D. L., Dekker, J., . . . de Vet, H. C. (2007). Quality criteria were proposed for measurement properties of health status questionnaires. *J Clin Epidemiol, 60*(1), 34-42.

Thomson Reuters. (2016). EndNote X7. New York: Thomson Reuters.

Tosevski, D. L., & Milovancevic, M. P. (2006). Stressful life events and physical health. *Current Opinion in Psychiatry, 19*(2), 184-189.

Trickey, D., Siddaway, A. P., Meiser-Stedman, R., Serpell, L., & Field, A. P. (2012). A meta-analysis of risk factors for post-traumatic stress disorder in children and adolescents. *Clinical Psychology Review, 32*(2), 122-138.

Twenge, J. M., & Nolen-Hoeksema, S. (2002). Age, gender, race, socioeconomic status, and birth cohort difference on the children's depression inventory: A meta-analysis. *Journal of abnormal psychology, 111*(4), 578.

Weaver, B., & Wuensch, K. L. (2013). SPSS and SAS programs for comparing Pearson correlations and OLS regression coefficients. *Behavior Research Methods, 45*(3), 880-895.

Wilkinson, L. (1999). Statistical methods in psychology journals: Guidelines and explanations. *American psychologist, 54*(8), 594.

Williamson, D. E., Birmaher, B., Anderson, B. P., Al-Shabbout, M., & Ryan, N. D. (1995). Stressful life events in depressed adolescents: the role of dependent events during the depressive episode. *Journal of the American Academy of Child & Adolescent Psychiatry, 34*(5), 591-598.

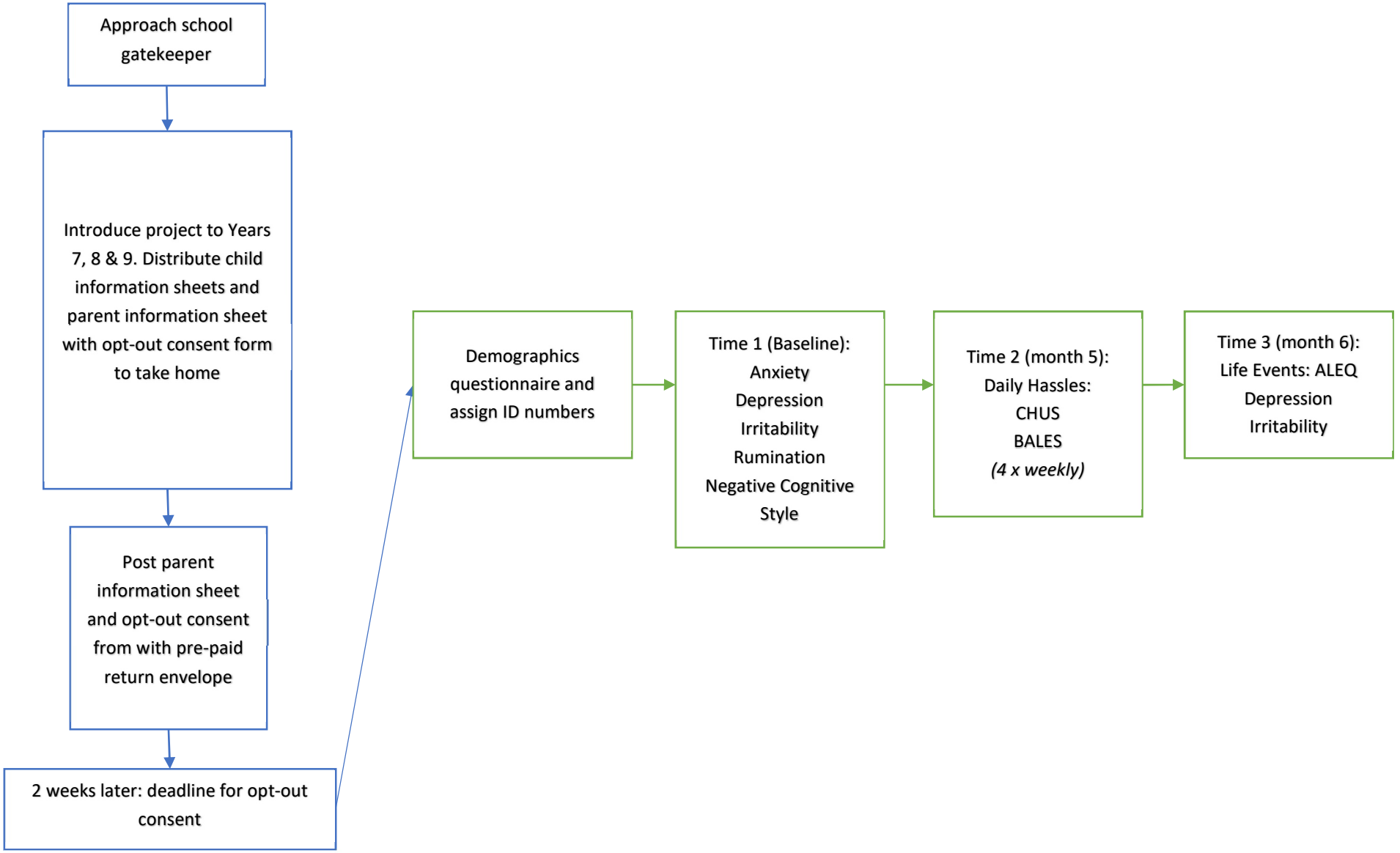
Williamson, D. E., Birmaher, B., Ryan, N. D., Shiffrin, T. P., Lusky, J. A., Protopapa, J., . . . Brent, D. A. (2003). The stressful life events schedule for children and adolescents: development and validation. *Psychiatry research, 119*(3), 225-241.

Wittchen, H.-U., Essau, C. A., Hecht, H., Teder, W., & Pfister, H. (1989). Reliability of life event assessments: test—retest reliability and fall-off effects of the Munich interview for the assessment of life events and conditions. *Journal of Affective Disorders, 16*(1), 77-91.

Wright, M., Creed, P., & Zimmer-Gembeck, M. J. (2010). The development and initial validation of a brief daily hassles scale suitable for use with adolescents. *European Journal of Psychological Assessment*, *26*, 220-226.

Appendices

Appendix 1. Flow Diagram of Study Design



CHUS Children's Hassles and Uplifts Scale, *BALES* Brief Adolescent Life Event Scale, *ALEQ* Adolescent Life Event Questionnaire

Appendix 2. Demographics at each time point

	T1	T2a	T2b	T2c	T2d	T3
N	249	193	195	179	186	196
Male : Female	130:102 (17 unknown)	106:86 (1 unknown)	109:86	95:84	110:76	108:87 (1 unknown)
Age in years (sd)	13.21 (.59)					13.72 (.61)
Ethnicity (%)						
Black	48 (51.0)	107 (55.4)	107 (54.9)	96 (53.6)	103 (55.4)	106 (54.1)
White	48 (19.3)	42 (21.8)	38 (19.5)	37 (20.7)	36 (19.4)	42 (21.4)
Asian	13 (5.2)	11 (5.7)	13 (6.7)	12 (6.7)	13 (7.0)	11 (5.6)
Mixed	30 (12.0)	23 (11.9)	26 (13.3)	24 (13.4)	24 (12.9)	28 (14.3)
Other	9 (3.6)	7 (3.6)	9 (4.6)	8 (4.5)	7 (3.8)	8 (4.1)
Unknown	31 (12.2)	3 (1.6)	2 (1.0)	2 (1.1)	3 (1.6)	1 (0.5)

Appendix 3. Children's Hassles and Uplifts Scale

Here is a list of things that young people sometimes feel bothered or upset about. We want to know if these things have happened to you ***during the last week***. Rate how often each thing has happened to you this week.

		Never 0	**A little** 1	**Some** 2	**A lot** 3
1	Someone at school teased you				
2	You had to clean your room				
3	You were punished for something you did not do				
4	You got punished for something you did wrong				
5	Your pet died				
6	Your best friend did not want to be your friend anymore				
7	Your mother, father or carer wasn't at home when you expected them				
8	You lost something				
9	Your mother, father or carer got sick				
10	Your mother, father or carer was angry at you for getting a bad school report				
11	Your teacher was angry at you because of your behaviour				
12	Your schoolwork was too hard				
13	You got into a fight with someone				
14	You didn't do well at sports				
15	You had to go to bed when you didn't feel like it				
16	Your mother, father or carer didn't have enough time to do something with you				
17	You didn't know the answer when the teacher asked you a question				
18	When your schoolmates were picking teams, you were one of the last ones to be picked				
19	Your mother, father or carers were fighting with each other				
20	Your mother, father or carer forgot to do something they said they would do				
21	You felt bored and wished there was something interesting to do				
22	Your brother or sisters annoyed you				
23	You didn't like the way you looked and wished you could be different (e.g. taller, stronger, better-looking)				
24	Someone did something better than you could				
25	You didn't have enough privacy (a time and place to be alone) when you wanted it				

Here is a list of things that children sometimes feel good about. We want to know if any of these things happened to you ***during the last week***. Rate how often each thing has happened to you this week.

		Never 0	A little 1	Some 2	A lot 3
1	You got a good mark at school				
2	You got a present you really wanted				
3	You won a game				
4	You found something you thought you'd lost				
5	You helped your sister or brother				
6	Your teacher was pleased with you				
7	You went out to eat				
8	Friends wanted you to be on their team				
9	There was a school trip				
10	You had a good time with your friends				
11	You gave a talk at school that went well				
12	Your parents/carers were pleased with your school marks				
13	You got some new clothes				
14	You did something special with your mum/dad/carer				
15	You had a good time at a party				
16	You were helped by your brother or sister				
17	Your played with your pet				
18	You did well at sports				
19	Your mother, father or carer spent time with you				
20	You made or fixed something by yourself				
21	You got a phone call, letter, text, email, message on social media				
22	You had fun joking with school mates				
23	You learned something new				
24	You made a new friend				
25	Your parents/carers agreed with you that something wasn't your fault.				

Appendix 4. Brief Adolescent Life Events Scale

Here is a list of things that young people say that they sometimes experience. Please rate how often each of these has happened to you ***in the last week***.

		Never 0	**A little** 1	**Some** 2	**A lot** 3
1	I argued with a family member				
2	I made up with a family member				
3	I got help from a family member when I needed it				
4	I did NOT get help from a family member when I needed it				
5	I was allowed to do something I wanted to				
6	I was NOT allowed to do something I wanted to				
7	I argued with a friend				
8	I made up with a friend				
9	I got help from a friend when I needed it				
10	I did NOT get help from a friend when I needed it				
11	A friend joined me for a special event when I asked				
12	A friend did NOT join me for a special event when I asked				
13	A classmate teased or threatened me				
14	A classmate defended me from others				
15	I was invited to join in with a group event				
16	I was excluded from a group event				
17	I had an enjoyable romantic date				
18	I had a disappointing romantic date				
19	I got a bad grade at school				
20	I got a good grade at school				
21	I completed an important assignment (on time)				
22	I did not complete an important assignment (or it was late)				
23	A teacher told me I did well on an assignment				
24	A teacher told me I did poorly on an assignment				
25	I discovered I can do something better than someone else				

		Never 0	**A little** 1	**Some** 2	**A lot** 3
26	I discovered I can NOT do something better than someone else				
27	I did something I felt embarrassed by				
28	I did something I felt proud of				
29	I did something outside of school that I was praised for				
30	I did something outside of school that I was criticised for				
31	My body changed in a way I wanted				
32	My body changed in a way I did NOT want				
33	I became sick or got injured				
34	I recovered from being sick or injured				
35	Someone insulted me because of the way I look				
36	Someone complimented me because of the way I look				

Appendix 5. Adolescent Life Event Questionnaire

INSTRUCTIONS: In this questionnaire we are interested in whether certain events have happened to you in the past 6 months. Please answer how often the following events have happened to you in the past **6 months** using this scale:

		Never 0	Rarely 1	Sometimes 2	Frequently 3	Always 4
Family and Parents						
1	Your parents divorced					
2	A close family member (parent, brother, sister) was hospitalized for serious injury/illness					
3	A close family member (parent, brother, sister) died.					
4	A close family member (parent, brother, sister) was arrested.					
5	You and your family moved to a new town, but you didn't want to move.					
6	You had an argument with a close family member (parent, brother, sister).					
7	A close family member (parent, brother, sister) lost their job.					
8	A close family member (parent, brother, sister) can't work due to injury/illness					
9	You had to do chores/ work you didn't want to do.					
10	You had to take care of brothers/ sisters when you didn't want to.					
11	You didn't spend as much time with close family members as you wanted to					
12	You couldn't seem to please your parents.					
13	You did something you didn't want to do to please a close family member.					
14	Your parents put you down.					
15	It seemed like your parents were disappointed with you					
16	A close family member had a significant medical or emotional					

	problem (e.g. heart disease, cancer, depression, etc.).					
17	You didn't receive the love, respect, or interest from parents that you wanted (e.g. parents didn't notice or compliment you on doing well).					
18	You fought with your parents over what you wanted for yourself, your future plan, or choice of friends					
19	Your parents forced you to achieve things you didn't want to do					
20	A close family member stopped giving love or affection from you.					
21	Your parents criticized you or yelled at you for not doing well in school					
22	Your parents grounded you.					
23	Your parents wouldn't let you go out with your friends					
24	You got in a fight with your parents over friends/ boyfriend/ girlfriend.					
Relationships						
25	A boyfriend/girlfriend broke up with you, but you still wanted to go out with them					
26	You didn't have a boyfriend/ girlfriend when you wanted one					
27	You got in a fight/ argument with a boyfriend/ girlfriend					
28	You couldn't seem to please girlfriend/ boyfriend when you wanted to					
29	Your girlfriend/ boyfriend criticized you					
30	You found out that boyfriend/ girlfriend had been cheating on you.					
31	You Did something to please your boyfriend/ girlfriend that you didn't want to do.					
School and classes						
32	You did poorly on, or failed, a test or class project.					

33	You did not have time to do well in school (example, working too many hours at work).					
34	You got a bad report card or end of term report.					
35	You didn't get to take a class you wanted to take.					
36	You didn't make a good set in school (e.g. top set for maths, English etc.)					
37	You had a bad teacher					
38	You didn't understand the material the teacher was teaching you					
39	You attended a class that you didn't like					
40	You didn't complete a required homework assignment for class					
41	You got in trouble with the teacher or principal					
42	You didn't get accepted for an extracurricular activity you wanted to be a part of					
Friends and social activities						
43	You didn't have as many friends as you would like to					
44	You weren't friends with the people you wanted to be friends with					
45	You didn't get invited to parties or dances					
46	You didn't have anyone to go out with on the weekends when you wanted to go out					
47	You had an argument with a close friend					
48	Your friends didn't seem to understand you					
49	You didn't have time to spend with your friends when you wanted to be with them					
50	You didn't talk or share feelings with your friends					
51	You got in a fight/ argument with your friends					

52	Your friends pressured you to do things you didn't want to do					
53	A close friend was arrested					
54	A close friend was hospitalized for a serious injury/illness					
55	A close friend died					
56	A close friend moved away					
57	A close friend stopped being so caring towards you					

Please list any other stressful, negative events that you can remember happening to you in the last **6 months**:

Appendix 6. Factor loadings from Principal Component Analysis for Children's Hassles and Uplifts Scale

		1	2
H1	Someone at school teased you	-0.074	0.334
H2	You had to clean your room	0.311	0.314
H3	You were punished for something you did not do	0.153	0.530
H4	You got punished for something you did wrong	0.152	0.529
H5	Your pet died	0.111	0.445
H6	Your best friend did not want to be your friend anymore	0.004	0.431
H7	Your mother, father or carer wasn't at home when you expected them	-0.055	0.509
H8	You lost something	0.249	0.474
H9	Your mother, father or carer got sick	0.168	0.643
H10	Your mother, father or carer was angry at you for getting a bad school report	0.096	0.453
H11	Your teacher was angry at you because of your behaviour	0.016	0.400
H12	Your schoolwork was too hard	-0.155	0.388
H13	You got into a fight with someone	0.252	0.522
H14	You didn't do well at sports	-0.109	0.212
H15	You had to go to bed when you didn't feel like it	0.070	0.457
H16	Your mother, father or carer didn't have enough time to do something with you	-0.219	0.412
H17	You didn't know the answer when the teacher asked you a question	-0.116	0.452
H18	When your schoolmates were picking teams, you were one of the last ones to be picked	0.073	0.500
H19	Your mother, father or carers were fighting with each other	-0.152	0.339
H20	Your mother, father or carer forgot to do something they said they would do	0.083	0.598
H21	You felt bored and wished there was something interesting to do	-0.083	0.520
H22	Your brother or sisters annoyed you	-0.102	0.474
H23	You didn't like the way you looked and wished you could be different (e.g. taller, stronger, better-looking)	-0.186	0.628
H24	Someone did something better than you could	-0.068	0.543
H25	You didn't have enough privacy (a time and place to be alone) when you wanted it	-0.212	0.564
U1	You got a good mark at school	0.381	-0.159
U2	You got a present you really wanted	0.645	0.023
U3	You won a game	0.520	0.023
U4	You found something you thought you'd lost	0.382	0.244
U5	You helped your sister or brother	0.532	-0.007
U6	Your teacher was pleased with you	0.582	-0.028
U7	You went out to eat	0.626	0.025
U8	Friends wanted you to be on their team	0.700	-0.045
U9	There was a school trip	0.401	0.387
U10	You had a good time with your friends	0.510	-0.143
U11	You gave a talk at school that went well	0.484	0.141
U12	Your parents/carers were pleased with your school marks	0.526	-0.240
U13	You got some new clothes	0.654	-0.025
U14	You did something special with your mum/dad/carer	0.702	-0.006
U15	You had a good time at a party	0.607	0.005
U16	You were helped by your brother or sister	0.611	-0.134

U17	Your played with your pet	0.124	0.120
U18	You did well at sports	0.600	0.045
U19	Your mother, father or carer spent time with you	0.669	-0.173
U20	You made or fixed something by yourself	0.578	-0.032
U21	You got a phone call, letter, text, email, message on social media	0.220	0.132
U22	You had fun joking with school mates	0.447	-0.073
U23	You learned something new	0.458	-0.051
U24	You made a new friend	0.485	0.186
U25	Your parents/carers agreed with you that something wasn't your fault.	0.574	0.052

Appendix 7. Factor loadings from Principal Components Analysis of the Brief Adolescent Life Events Scale

		1	2	Original coding
1	I argued with a family member	-0.068	0.585	neg
2	I made up with a family member	0.277	0.393	pos
3	I got help from a family member when I needed it	0.540	0.004	pos
4	I did NOT get help from a family member when I needed it	-0.067	0.475	neg
5	I was allowed to do something I wanted to	0.389	-0.147	pos
6	I was NOT allowed to do something I wanted to	0.133	0.519	neg
7	I argued with a friend	0.084	0.642	neg
8	I made up with a friend	0.348	0.442	pos
9	I got help from a friend when I needed it	0.446	0.032	pos
10	I did NOT get help from a friend when I needed it	0.149	0.491	neg
11	A friend joined me for a special event when I asked	0.509	0.236	pos
12	A friend did NOT join me for a special event when I asked	0.052	0.355	neg
13	A classmate teased or threatened me	0.000	0.598	neg
14	A classmate defended me from others	0.476	0.258	pos
15	I was invited to join in with a group event	0.609	0.070	pos
16	I was excluded from a group event	0.057	0.368	neg
17	I had an enjoyable romantic date	0.094	0.366	pos
18	I had a disappointing romantic date	-0.011	0.328	neg
19	I got a bad grade at school	-0.153	0.462	neg
20	I got a good grade at school	0.615	-0.086	pos
21	I completed an important assignment (on time)	0.629	-0.079	pos
22	I did not complete an important assignment (or it was late)	0.004	0.504	neg
23	A teacher told me I did well on an assignment	0.700	-0.034	pos
24	A teacher told me I did poorly on an assignment	-0.052	0.537	neg
25	I discovered I can do something better than someone else	0.650	0.156	pos

26	I discovered I can NOT do something better than someone else	0.041	0.359	neg
27	I did something I felt embarrassed by	0.062	0.603	neg
28	I did something I felt proud of	0.711	-0.013	pos
29	I did something outside of school that I was praised for	0.724	-0.067	pos
30	I did something outside of school that I was criticised for	0.173	0.424	neg
31	My body changed in a way I wanted	0.488	0.150	pos
32	My body changed in a way I did NOT want	0.020	0.418	neg
33	I became sick or got injured	0.342	0.197	neg
34	I recovered from being sick or injured	0.399	0.280	pos
35	Someone insulted me because of the way I look	0.042	0.530	neg
36	Someone complimented me because of the way I look	0.960	0.062	pos

Appendix 8. Factor loadings from Principal Components Analysis of the Adolescent Life Events Questionnaire

		1	2	3
1	Your parents divorced	0.26	0.18	0.02
2	A close family member (parent, brother, sister) was hospitalized for serious injury/illness	0.07	-0.09	0.43
3	A close family member (parent, brother, sister) died.	0.06	0.10	0.41
4	A close family member (parent, brother, sister) was arrested.	0.09	0.46	0.01
5	You and your family moved to a new town, but you didn't want to move.	-0.02	0.17	0.26
6	You had an argument with a close family member (parent, brother, sister).	0.07	0.09	0.63
7	A close family member (parent, brother, sister) lost their job.	0.21	-0.01	0.11
8	A close family member (parent, brother, sister) can't work due to injury/illness	-0.02	0.01	0.30
9	You had to do chores/ work you didn't want to do.	0.16	0.27	0.45
10	You had to take care of brothers/ sisters when you didn't want to.	0.39	0.45	0.19
11	You didn't spend as much time with close family members as you wanted to	0.02	0.55	0.36
12	You couldn't seem to please your parents.	0.27	0.47	0.34
13	You did something you didn't want to do to please a close family member.	-0.04	0.46	0.36
14	Your parents put you down.	0.26	0.67	0.06
15	It seemed like your parents were disappointed with you	0.29	0.49	0.38
16	A close family member had a significant medical or emotional problem	-0.01	-0.08	0.43
17	You didn't receive the love, respect, or interest from parents that you wanted	0.03	0.60	0.18
18	You fought with your parents over what you wanted for yourself, your future plan, or choice of friends	0.36	0.39	0.44
19	Your parents forced you to achieve things you didn't want to do	0.00	0.44	0.28
20	A close family member stopped giving love or affection from you.	-0.05	0.61	0.07
21	Your parents criticized you or yelled at you for not doing well in school	0.12	0.27	0.52
22	Your parents grounded you.	-0.02	0.39	0.43
23	Your parents wouldn't let you go out with your friends	-0.04	0.43	0.40
24	You got in a fight with your parents over friends/ boyfriend/ girlfriend.	0.14	0.12	0.53
25	A boyfriend/girlfriend broke up with you, but you still wanted to go out with them	0.24	0.58	0.04
26	You didn't have a boyfriend/ girlfriend when you wanted one	0.33	0.60	0.09
27	You got in a fight/ argument with a boyfriend/ girlfriend	0.22	0.51	0.24
28	You couldn't seem to please girlfriend/ boyfriend when you wanted to	0.81	0.31	-0.02
29	Your girlfriend/ boyfriend criticized you	0.68	0.34	-0.01
30	You found out that boyfriend/ girlfriend had been cheating on you.	0.06	0.60	-0.03

31	You Did something to please your boyfriend/ girlfriend that you didn't want to do.	0.78	0.15	0.02
32	You did poorly on, or failed, a test or class project.	0.02	0.44	0.46
33	You did not have time to do well in school (example, working too many hours at work).	-0.03	0.36	0.12
34	You got a bad report card or end of term report.	0.48	0.18	0.38
35	You didn't get to take a class you wanted to take.	0.02	0.25	0.23
36	You didn't make a good set in school (e.g. top set for maths, English etc.)	0.42	0.55	0.02
37	You had a bad teacher	0.29	0.04	0.51
38	You didn't understand the material the teacher was teaching you	0.11	0.56	0.32
39	You attended a class that you didn't like	0.23	0.04	0.55
40	You didn't complete a required homework assignment for class	0.17	0.26	0.43
41	You got in trouble with the teacher or principal	0.26	0.13	0.55
42	You didn't get accepted for an extracurricular activity you wanted to be a part of	0.12	0.46	-0.06
43	You didn't have as many friends as you would like to	0.47	0.38	0.06
44	You weren't friends with the people you wanted to be friends with	0.62	0.48	0.01
45	You didn't get invited to parties or dances	0.52	0.39	0.01
46	You didn't have anyone to go out with on the weekends when you wanted to go out	0.44	0.60	0.00
47	You had an argument with a close friend	0.51	-0.02	0.28
48	Your friends didn't seem to understand you	0.59	0.27	0.16
49	You didn't have time to spend with your friends when you wanted to be with them	0.43	0.51	0.27
50	You didn't talk or share feelings with your friends	0.46	0.43	0.20
51	You got in a fight/ argument with your friends	0.56	0.18	0.28
52	Your friends pressured you to do things you didn't want to do	0.58	0.55	0.03
53	A close friend was arrested	0.75	-0.02	0.07
54	A close friend was hospitalized for a serious injury/illness	0.61	-0.14	0.07
55	A close friend died	0.79	-0.08	0.02
56	A close friend moved away	0.55	-0.08	0.17
57	A close friend stopped being so caring towards you	0.68	0.20	0.11